

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
10 May 2001 (10.05.2001)

PCT

(10) International Publication Number
WO 01/32882 A2

(51) International Patent Classification⁷: C12N 15/31,
C12Q 1/68, C12N 1/21, C07K 14/315, 16/12, A61K
39/09, 48/00, G01N 33/53, 33/68

of Pathology, Tennis Court Road, Cambridge CB2 1QP
(GB). WELLS, Jeremy, Mark [GB/GB]; Institute of
Food Research, Norwich Laboratory, Norwich Research
Park, Colney, Norwich NR4 7UA (GB). HANNIFFY,
Sean, Bosco [IE/GB]; University of Cambridge, Dept. of
Pathology, Tennis Court Road, Cambridge CB2 1QP (GB).

(21) International Application Number: PCT/GB00/03437

(22) International Filing Date:
7 September 2000 (07.09.2000) (74) Agents: CHAPMAN, Paul, William et al.; Kilburn &
Strode, 20 Red Lion Street, London WC1R 4PJ (GB).

(25) Filing Language:

English

(81) Designated States (*national*): CA, CN, JP, US.

(26) Publication Language:

English

(84) Designated States (*regional*): European patent (AT, BE,
CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC,
NL, PT, SE).

(30) Priority Data:

9921125.2 7 September 1999 (07.09.1999) GB

Published:

— *Without international search report and to be republished
upon receipt of that report.*

(71) Applicant (*for all designated States except US*): MICRO-
BIAL TECHNIKS LIMITED [GB/GB]; 20 Trumpington
Street, Cambridge CB2 1QA (GB).

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): LE PAGE, Richard,
William, Falla [GB/GB]; University of Cambridge, Dept.

WO 01/32882 A2

(54) Title: NUCLEIC ACIDS AND PROTEINS FROM GROUP B STREPTOCOCCUS

(57) Abstract: Novel protein antigens from Group B Streptococcus are described, together with the nucleic acid sequences encoding them. The use of vaccines and screening methods is also described.

Proteins

The present invention relates to proteins derived from *Streptococcus agalactiae*, nucleic acid molecules encoding such proteins, and the use of the proteins as antigens and/or immunogens and in detection/diagnosis. It also relates to a method for the rapid screening of bacterial genomes to isolate and characterise bacterial cell envelope associated or secreted proteins.

The Group B Streptococcus (GBS) (*Streptococcus agalactiae*) is an encapsulated bacterium which emerged in the 1970s as a major pathogen of humans causing sepsis and meningitis in neonates as well as adults. The incidence of early onset neonatal infection during the first 5 days of life varies from 0.7 to 3.7 per 1000 live births and causes mortality in about 20% of cases. Between 25-50% of neonates surviving early onset infections frequently suffer neurological sequelae. Late onset neonatal infections occur from 6 days to three months of age at a rate of about 0.5 - 1.0 per 1000 live births.

There is an established association between the colonisation of the maternal genital tract by GBS at the time of birth and the risk of neonatal sepsis. In humans it has been established that the rectum may act as a reservoir for GBS. Susceptibility in the neonate is correlated with the a low concentration or absence of IgG antibodies to the capsular polysaccharides found on GBS causing human disease. In the USA strains isolated from clinical cases usually belong to capsular serotypes Ia, Ib, II, III although serotype V may be of increasing significance. Type VIII GBS is the major cause of neonatal sepsis in Japan.

A possible means of prevention involves intra or postpartum administration of antibiotics to the mother but there are concerns that this might lead to the emergence

of resistant organisms and in some cases allergic reactions. Vaccination of the adolescent females to induce long lasting maternally derived immunity is one of the most promising approaches to prevent GBS infections in neonates. The capsular polysaccharide antigens of these organisms have attracted most attention as with regard to vaccine development. Studies in healthy adult volunteers have shown that serotype Ia, II and III polysaccharides are non-toxic and immunogenic in approximately 65%, 95% and 70% of non-immune adults respectively. One of the problems with using capsule antigens as vaccines is that the response rates vary according to pre-immunisation status and the polysaccharide antigen and not all vaccinees produce adequate levels of IgG antibody as indicated in vaccination studies with GBS polysaccharides in human volunteers.

Some people do not respond despite repeated stimuli. These properties are due to the T-independent nature of polysaccharide antigens. One strategy to enhance the immunogenicity of these vaccines is to enhance the T cell dependent properties of polysaccharides by conjugating them to a protein. The use of polysaccharide conjugates looks promising but there are still unresolved questions concerning the nature of the carrier protein. A conjugate vaccine against GBS would require at least 4 different conjugates to be prepared adding to the cost of a vaccine.

Approaches to vaccination against GBS infections which rely on the use of capsular polysaccharides have the disadvantage that response rates are likely to vary considerably according to pre-immunisation status and the particular type of polysaccharide antigen used. Results of trials with conjugate vaccines in human volunteers have indicated that response rates may only be around 65% for some of the key capsule antigens (Larsson *et al.*, *Infection and Immunity* **64**:3518-3523 (1996)). It is also not clear whether all individuals responding to the vaccine would have adequate levels of polysaccharide specific IgG which can cross the placenta and

afford immunity to neonates. By conjugating a protein carrier to the polysaccharide antigen it may be possible to convert them to T-cell dependent antigens and enhance their immunogenicity.

5 Preliminary studies with GBS type III polysaccharide-tetanus toxoid conjugate have been encouraging (Baker *et al.*, *Reviews of Infectious Diseases* 7:458-467 (1985), Baker *et al.*, *The New England Journal of Medicine* 319:1180-1185 (1988), Paoletti *et al.*, *Infection and Immunity* 64:677-679 (1996), Paoletti *et al.*, *Infection and Immunity* 62:3236-3243 (1994)) but in developed countries the use of tetanus may be
10 disadvantageous since most adults will have been immunised against tetanus within the past five years. Additional boosters with tetanus toxoid may cause adverse reactions (Boyer., *Current Opinions in Pediatrics* 7:13-18 (1995)). The polysaccharide conjugate vaccines have the disadvantage of being costly to produce and manufacture in comparison with many other kinds of vaccines. There is also the
15 possible risk of problems caused by the cross reactivity between GBS polysaccharides and sialic acid-containing human glycoproteins.

Recent evidence suggests that bacterial surface proteins also may be useful to confer immunity. A protein called Rib which is found on most serotype III strains but rarely
20 on serotypes Ia, Ib or II confers immunity to challenge with Rib expressing GBS in animal models (Stalhammar-Carlemalm *et al.*, *Journal of Experimental Medicine* 177:1593-1603 (1993)). Another surface protein of interest as a component of a vaccine is the alpha antigen of the C proteins which protected vaccinated mice against lethal infection with strains expressing alpha protein. The amount of this
25 antigen expressed by GBS strains varies markedly, however an alternative to polysaccharides as antigens is the use of protein antigens derived from GBS. Recent evidence suggest that the GBS surface associated proteins Rib and alpha C protein may be used to confer immunity to GBS infections in experimental model systems

(Stalhammar-Carlemalm *et al.*, (1993) [*supra*], Larsson *et al.*, (1996) [*supra*]). However these two proteins are not conserved in all serotypes of GBS which cause disease in humans. Assuming that these antigens would be immunogenic and elicit protective level responses in humans they would not confer protection against all 5 infections caused by GBS as 10% of infectious Group B streptococci do not express Rib or C protein alpha.

This invention seeks to overcome the problem of vaccination against GBS by using a novel screening method specifically designed to identify those Group B 10 Streptococcus genes encoding bacterial cell surface associated or secreted proteins. The proteins expressed by these genes may be immunogenic, and therefore may be useful in the prevention and treatment of Group B Streptococcus infection. For the purposes of this application, the term immunogenic means that these proteins will elicit a protective immune response within a subject. Using this novel screening 15 method a number of genes encoding novel Group B Streptococcus proteins have been identified.

Thus in a first aspect, the present invention provides a Group B Streptococcus protein, polypeptide or peptide having a sequence selected from those shown in 20 figure 1, or fragments or derivatives thereof.

It will be apparent to the skilled person that proteins and polypeptides included within this group may be cell surface receptors, adhesion molecules, transport proteins, membrane structural proteins, and/or signalling molecules.

25 Alterations in the amino acid sequence of a protein can occur which do not affect the function of a protein. These include amino acid deletions, insertions and substitutions and can result from alternative splicing and/or the presence of multiple translation

start sites and stop sites. Polymorphisms may arise as a result of the infidelity of the translation process. Thus changes in amino acid sequence may be tolerated which do not affect the protein's function.

5 Thus, the present invention includes derivatives or variants of the proteins, polypeptides, and peptides of the present invention which show at least 50% identity to the proteins, polypeptides and peptides described herein. Preferably the degree of sequence identity is at least 60% and preferably it is above 75%. More preferably still it is above 80%, 90% or even 95%.

10

The term identity can be used to describe the similarity between two polypeptide sequences. A software package well known in the art for carrying out this procedure is the CLUSTAL program. It compares the amino acid sequences of two polypeptides and finds the optimal alignment by inserting spaces in either sequence as appropriate. The amino acid identity or similarity (identity plus conservation of amino acid type) for an optimal alignment can also be calculated using a software package such as BLASTx. This program aligns the largest stretch of similar sequence and assigns a value to the fit. For any one pattern comparison several regions of similarity may be found, each having a different score. One skilled in the art will appreciate that two polypeptides of different lengths may be compared over the entire length of the longer fragment. Alternatively small regions may be compared. Normally sequences of the same length are compared for a useful comparison to be made.

20

25 Manipulation of the DNA encoding the protein is a particularly powerful technique for both modifying proteins and for generating large quantities of protein for purification purposes. This may involve the use of PCR techniques to amplify a desired nucleic acid sequence. Thus the sequence data provided herein can be used to

design primers for use in PCR so that a desired sequence can be targeted and then amplified to a high degree.

5 Typically primers will be at least five nucleotides long and will generally be at least ten nucleotides long (e.g. fifteen to twenty-five nucleotides long). In some cases primers of at least thirty or at least thirty-five nucleotides in length may be used.

10 As a further alternative chemical synthesis may be used. This may be automated. Relatively short sequences may be chemically synthesised and ligated together to provide a longer sequence.

Thus in a further aspect, the present invention provides, a nucleic acid molecule comprising or consisting of a sequence which is:

- 15 (i) any of the DNA sequences set out in figure 1 herein or their RNA equivalents;
- (ii) a sequence which is complementary to any of the sequences of (i);
- (iii) a sequence which codes for the same protein or polypeptide, as those sequences of (i) or (ii);
- (iv) a sequence which shows substantial identity with any of those of (i),
20 (ii) and (iii); or
- (v) a sequence which codes for a derivative or fragment of a nucleic acid molecule shown in figure 1.

The term identity can also be used to describe the similarity between two individual DNA sequences. The 'bestfit' program (Smith and Waterman, *Advances in applied Mathematics*, 482-489 (1981)) is one example of a type of computer software used to find the best segment of similarity between two nucleic acid sequences, whilst the GAP program enables sequences to be aligned along their whole length and finds the optimal alignment by inserting spaces in either sequence as appropriate.

The present invention includes nucleic acid sequences which show at least 50% identity to the nucleic acid sequences described herein. Preferably the degree of sequence identity is at least 60% and preferably it is above 75%. More preferably still it is above 80%, 90% or even 95%.

5

The term 'RNA equivalent' when used above indicates that a given RNA molecule has a sequence which is complementary to that of a given DNA molecule, allowing for the fact that in RNA 'U' replaces 'T' in the genetic code. The nucleic acid molecule may be in isolated, recombinant or chemically synthetic form.

10

DNA constructs can readily be generated using methods well known in the art. These techniques are disclosed, for example in J. Sambrook *et al, Molecular Cloning 2nd Edition*, Cold Spring Harbour Laboratory Press (1989). Modifications of DNA constructs and the proteins expressed such as the addition of promoters, enhancers, signal sequences, leader sequences, translation start and stop signals and DNA stability controlling regions, or the addition of fusion partners may then be facilitated.

20

Normally the DNA construct will be inserted into a vector which may be any suitable vector, including plasmid, virus, bacteriophage, transposon, minichromosome, liposome or mechanical carrier. The expression vectors of the invention are DNA constructs suitable for expressing DNA which encodes the desired protein product which may include: (a) a regulatory element (e.g. a promoter, operator, activator, repressor and/or enhancer), (b) a structural or coding sequence which is transcribed into mRNA and (c) appropriate transcription, translation, initiation and termination sequences. The vector may further comprise a selectable marker, for example antibiotic resistance, which facilitates the selection and/or identification of cells containing the vector.

Expression of the protein is achieved by the transformation or transfection of the vector into a host cell which may be of eukaryotic or prokaryotic origin. For the production of recombinant protein, expression may be inducible expression or 5 expression only in certain types of cells or both inducible and cell-specific. Particularly preferred among inducible vectors are vectors that can be induced for expression by environmental factors that are easy to manipulate, such as temperature and nutrient additives. A variety of suitable vectors, including constitutive and inducible expression vectors for use in prokaryotic and eukaryotic 10 hosts, are well known and employed routinely by those skilled in the art.

A great variety of expression vectors can be used to express the Group B Streptococcus protein(s) of the invention. Such vectors include, among others, chromosomal, episomal and virus-derived vectors, for example, vectors derived 15 from bacterial plasmids, from bacteriophage, from transposons, from yeast elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids, all may be used in accordance with the invention. Generally, any 20 vector suitable to maintain, propagate or express nucleic acid to express a polypeptide in a host may be used for expression in this regard. Such vectors thus form yet a further aspect of the invention.

The appropriate DNA sequence may be inserted into the vector by any of a variety 25 of well-known and routine techniques.

The nucleic acid sequence in the expression vector is operatively linked to appropriate expression control sequence(s) including, for instance, a promoter to

direct mRNA transcription. Representatives of such promoters include, but are not limited to, the phage lambda PL promoter, the T3 and T7 promoters, the *E.coli* lac, trp, tac, and λ P_L promoters, the microbial eukaryote GAL, glucoamylase and cellobiohydrolase promoters and the mammalian metallothionein (mouse) and heat-shock (human) promoters.

In general, expression vectors will contain sites for transcription initiation and termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of mature transcripts expressed by the constructs will generally include a translation initiating AUG at the beginning and a termination codon appropriately positioned at the end of the polypeptide to be translated.

Representative examples of appropriate hosts for recombinant expression of the Group B Streptococcus protein(s) of the invention include bacterial cells, such as *streptococci*, *staphylococci*, *E.coli*, *streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa and Bowes melanoma cells; and plant cells. Such host cells form yet a further aspect of the present invention.

20

Microbial cells employed in the expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agent, such methods which are known to those skilled in the art.

25

The polypeptide can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulphate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose,

chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Well known techniques for refolding protein may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

5

The Group B Streptococcus proteins described herein can additionally be used as target antigens to raise antibodies, or to generate affibodies. These can be used to detect Group B Streptococcus.

- 10 Thus in a further aspect the present invention provides, an antibody, affibody, or a derivative thereof which binds to any one or more of the proteins, polypeptides, peptides, fragments or derivatives thereof, as described herein.

- 15 Antibodies within the scope of the present invention may be monoclonal or polyclonal. Polyclonal antibodies can be raised by stimulating their production in a suitable animal host (e.g. a mouse, rat, guinea pig, rabbit, sheep, goat or monkey) when a protein as described herein, or a homologue, derivative or fragment thereof, is injected into the animal. If desired, an adjuvant may be administered together with the protein. Well-known adjuvants include Freund's adjuvant (complete and incomplete) and aluminium hydroxide. The antibodies can then be purified by virtue of their binding to a protein as described herein and by many other means well-known to those skilled in the art.

- 20 Monoclonal antibodies can be produced from hybridomas. These can be formed by fusing myeloma cells and spleen cells which produce the desired antibody in order to form an immortal cell line. Thus the well-known Kohler & Milstein technique (*Nature* 256 (1975)) or subsequent variations upon this technique can be used.

Techniques for producing monoclonal and polyclonal antibodies that bind to a particular polypeptide/protein are now well developed in the art. They are discussed in standard immunology textbooks, for example in Roitt *et al*, *Immunology* second edition (1989), Churchill Livingstone, London.

5

In addition to whole antibodies, the present invention includes derivatives thereof which are capable of binding to proteins etc as described herein. Thus the present invention includes antibody fragments and synthetic constructs. Examples of antibody fragments and synthetic constructs are given by Dougall *et al*., *Tibtech* 12 372-379 (September 10 1994).

15

Antibody fragments include, for example, Fab, F(ab')₂ and Fv fragments. Fv fragments can be modified to produce a synthetic construct known as a single chain Fv (scFv) molecule. This includes a peptide linker covalently joining V_b and V_i regions, which contributes to the stability of the molecule. Other synthetic constructs that can be used include CDR peptides. These are synthetic peptides comprising antigen-binding determinants. Peptide mimetics may also be used. These molecules are usually conformationally restricted organic rings that mimic the structure of a CDR loop and that include antigen-interactive side chains.

20

Synthetic constructs include chimaeric molecules. Thus, for example, humanised (or primatised) antibodies or derivatives thereof are within the scope of the present invention. An example of a humanised antibody is an antibody having human framework regions, but rodent hypervariable regions. Ways of producing chimaeric 25 antibodies are discussed for example by Morrison *et al* in *PNAS*, 81, 6851-6855 (1984) and by Takeda *et al* in *Nature*, 314, 452-454 (1985).

Synthetic constructs also include molecules comprising an additional moiety that provides the molecule with some desirable property in addition to antigen binding. For example the moiety may be a label (e.g. a fluorescent or radioactive label). Alternatively, it may be a pharmaceutically active agent.

5

Affibodies are proteins which are found to bind to target proteins with a low dissociation constant. They are selected from phage display libraries expressing a segment of the target protein of interest (Nord K, Gunnarsson E, Ringdahl J, Stahl S, Uhlen M, Nygren PA, Department of Biochemistry and Biotechnology, Royal Institute 10 of Technology (KTH), Stockholm, Sweden).

In a further aspect the invention provides an immunogenic composition comprising one or more proteins, polypeptides, peptides, fragments or derivatives thereof, or nucleotide sequences described herein. The immunogenic composition may include 15 nucleic acid sequences ID-65 and/or ID-66 as described herein. Alternatively, the immunogenic composition may comprise proteins/polypeptides including ID-65, ID-83, ID-89, ID-93 and/or ID-96 as described herein, or fragments or derivatives thereof. A composition of this sort may be useful in the treatment or prevention of Group B Streptococcus infection in subject. In a preferred aspect of the invention the 20 immunogenic composition is a vaccine.

In other aspects the invention provides:

i) Use of an immunogenic composition as described herein in the preparation of 25 a medicament for the treatment or prophylaxis of Group B Streptococcus infection. Preferably the medicament is a vaccine.

- ii) A method of detection of Group B Streptococcus which comprises the step of bringing into contact a sample to be tested with at least one antibody, affibody, or a derivative thereof, as described herein.
- 5 iii) A method of detection of Group B Streptococcus which comprises the step of bringing into contact a sample to be tested with at least one protein, polypeptide, peptide, fragments or derivatives as described herein.
- iv) A method of detection of Group B Streptococcus which comprises the step of bringing into contact a sample to be tested with at least one nucleic acid molecule as described herein.
- 10 v) A kit for the detection of Group B Streptococcus comprising at least one antibody, affibody, or derivatives thereof, described herein.
- 15 vi) A kit for the detection of Group B Streptococcus comprising at least one Group B Streptococcus protein, polypeptide, peptide, fragment or derivative thereof, as described herein.
- 20 vii) A kit for the detection of Group B Streptococcus comprising at least one nucleic acid of the invention.

As described previously, the novel proteins described herein are identified and isolated using a screening method which specifically identifies those Group B Streptococcus genes encoding bacterial cell envelope associated or secreted proteins.

Given that the inventors have identified a group of important proteins, such proteins are potential targets for anti-microbial therapy. It is necessary, however, to

determine whether each individual protein is essential for the organism's viability. Thus, the present invention also provides a method of determining whether a protein or polypeptide as described herein represents a potential anti-microbial target which comprises inactivating said protein and determining whether Group B Streptococcus
5 is still viable.

A suitable method for inactivating the protein is to effect selected gene knockouts, ie prevent expression of the protein and determine whether this results in a lethal change. Suitable methods for carrying out such gene knockouts are described in Li
10 *et al*, *P.N.A.S.*, 94:13251-13256 (1997) and Kolkman *et al.*, *Journal of Biological Chemistry* 272: 19502-19508 (1997); Kolkman *et al.*, *Journal of Bacteriology* 178:
3736-3741 (1996).

In a final aspect the present invention provides the use of an agent capable of antagonising, inhibiting or otherwise interfering with the function or expression of a
15 protein or polypeptide of the invention in the manufacture of a medicament for use in the treatment or prophylaxis of Group B Streptococcus infection.

The invention will now be described by means of the following examples which should not in any way be construed as limiting. The examples refer to the figures in
20 which:

Fig 1: (A) Shows a number of full length nucleotide sequences encoding antigenic Group B Streptococcus proteins and the corresponding amino acid sequences.

25

Fig 2: Shows the results of vaccine trials using the proteins ID-65 and ID-66;

Fig 3: Shows a number of oligonucleotide primers used in the screening process

nucS1 primer designed to amplify a mature form of the nuc A gene

nucS2- primer designed to amplify a mature form of the nuc A gene.

5 nucS3 primer designed to amplify a mature form of the nuc A gene

nucR primer designed to amplify a mature form of the nuc A gene

nucseq primer designed to sequence DNA cloned into the pTREP-Nuc vector

pTREPF nucleic acid sequence containing recognition site for ECORV. Used for cloning fragments into pTREX7.

10 pTREPR nucleic acid sequence containing recognition site for BAMH1.

Used for cloning fragments into pTREX7.

PUCF forward sequencing primer, enables direct sequencing of cloned DNA fragments.

15 VR example of gene specific primer used to obtain further antigen DNA sequence by the method of DNA walking.

V1 example of gene specific primer used to obtain further antigen DNA sequence by the method of DNA walking.

V2 example of gene specific primer used to obtain further antigen DNA sequence by the method of DNA walking.

20

Fig 4: (i) Schematic presentation of the nucleotide sequence of the unique gene cloning site immediately upstream of the mature *nuc* gene in pTREP1-*nuc1*, pTREP1-*nuc2* and pTREP1-*nuc3*. Each of the pTREP-*nuc* vectors contain an EcoRV (a SmaI site in pTREP1-*nuc2*) cleavage site which allows cloning of genomic DNA fragments in 3 different frames with respect to the mature *nuc* gene.

25

(ii) A physical and genetic summary map of the pTREP1-*nuc* vectors. The expression cassette incorporating *nuc*, the macrolides, lincosamides and

streptogramin B (MLS) resistance determinant, and the replicon (rep) *Ori-pAMβ1* are depicted (not drawn to scale).

5 (iii) Schematic presentation of the expression cassette showing the various sequence elements involved in gene expression and location of unique restriction endonuclease sites (not drawn to scale).

10

Fig 5: SDS-PAGE analysis of a purified preparation of the His-tagged ID-65 and ID-83 protein antigens (predicted molecular weights of 57,144 and 25,000 daltons respectively) on a 12% polyacrylamide gel. Lanes: MW, molecular weight standards; 1, His-tagged ID-65 protein; 2, His-tagged ID-83 protein

15

Fig 6: SDS PAGE analysis of a purified preparation of the His-tagged ID-93 protein antigen (predicted molecular weight = 28,000 daltons) on a 12% polyacrylamide gel.

Lanes: MW, molecular weight standards; 1, His-tagged ID-93 protein.

20

Fig 7: SDS PAGE analysis of a purified preparation of the His-tagged ID-89 and ID-96 protein antigens (predicted molecular weights of 35,000 and 31,000 daltons respectively) on a 12% polyacrylamide gel.

Lanes: MW, molecular weight standards; 1, His-tagged ID-89 protein; 2, His-tagged ID-96 protein.

25

Fig 8: IgG Titres against the ID-65 and ID-83 proteins

1 = ID-65 + Alum Group - Bleed at 5 weeks

2 = PBS + Alum Control Group - Bleed at 5 weeks

(For groups 1 and 2, ELISAs were performed on purified ID-65 protein)

3 = ID-83 + Alum Group - Bleed at 5 weeks

17.

4 = PBS + Alum Control Group - Bleed at 5 weeks

(For groups 3 and 4, ELISAs were performed on purified ID-83 protein)

Fig 9: Shows the results of vaccine trials using the protein ID-93.

5

Fig 10: IgG titres against the ID-93 protein.

1 = ID-93 + Alum Group - Bleed at 3 weeks

2 = ID-93 + Alum Group - Bleed at 6 weeks

3 = PBS + Alum Control Group - Bleed at 3 weeks

10 4 = PBS + Alum Control Group - Bleed at 6 weeks

Fig 11: IgG titres against the ID-89 and ID-96 proteins

1 = ID-89 + TitreMax Gold Group - Bleed at 3 weeks

2 = ID-89 + TitreMax Gold - Bleed at 6 weeks

15 3 = PBS + TitreMax Gold Control Group - Bleed at 3 weeks

4 = PBS + TitreMax Gold Control Group - Bleed at 6 weeks

5 = ID-96 + TitreMax Gold Group - Bleed at 3 weeks

6 = ID-96 + TitreMax Gold Group - Bleed at 6 weeks

7 = PBS + TitreMax Gold Control Group - Bleed at 3 weeks

20 8 = PBS + TitreMax Gold Control Group - Bleed at 6 weeks

For Groups 1-4, ELISAs were performed on purified ID-89 protein.

For Groups 5-6, ELISAs were performed on purified ID-96 protein.

25

Fig 12: Southern blot analysis of genomic DNA. Genomic DNA from each of the strains listed in Table 7 was digested completely with *Hin* DIII (NEB) and electrophoresed at 40 Volts for 6 hours in 0.8% agarose, transferred onto Hybond N⁺ (Amersham) membrane by Southern blot and hybridised with the

digoxigenin-labelled *rib* gene probe. Specifically bound DNA probe was identified using the DIG Nucleic Acid Detection Kit (Boehringer Mannheim).

Fig 13: Southern blot analysis of genomic DNA. Genomic DNA from each of the strains listed in Table 6 was digested completely with *Hin* DIII (NEB) and electrophoresed at 40 Volts for 6 hours in 0.8% agarose, transferred onto Hybond N⁺ (Amersham) membrane by Southern blot and hybridised with the digoxigenin-labelled ID-65 gene probe. Specifically bound DNA probe was identified using the DIG Nucleic Acid Detection Kit (Boehringer Mannheim).

Fig 14: Southern blot analysis of genomic DNA. Genomic DNA from each of the strains listed in Table 6 was digested completely with *Hin* DIII (NEB) and electrophoresed at 40 Volts for 6 hours in 0.8% agarose, transferred onto Hybond N⁺ (Amersham) membrane by Southern blot and hybridised with the digoxigenin-labelled ID-89 gene probe. Specifically bound DNA probe was identified using the DIG Nucleic Acid Detection Kit (Boehringer Mannheim).

Fig 15: Southern blot analysis of genomic DNA. Genomic DNA from each of the strains listed in Table 6 was digested completely with *Hin* DIII (NEB) and electrophoresed at 40 Volts for 6 hours in 0.8% agarose, transferred onto Hybond N⁺ (Amersham) membrane by Southern blot and hybridised with the digoxigenin-labelled ID-93 gene probe. Specifically bound DNA probe was identified using the DIG Nucleic Acid Detection Kit (Boehringer Mannheim).

Fig 16: Southern blot analysis of genomic DNA. Genomic DNA from each of the strains listed in Table 6 was digested completely with *Eco* RI (NEB) and electrophoresed at 40 Volts for 6 hours in 0.8% agarose, transferred onto Hybond N⁺ (Amersham) membrane by Southern blot and hybridised with the

digoxigenin-labelled ID-96 gene probe. Specifically bound DNA probe was identified using the DIG Nucleic Acid Detection Kit (Boehringer Mannheim).

5

Example 1

Gene/partial gene sequences putatively encoding exported proteins in *S. agalactiae* have been identified, unless stated otherwise, using the nuclease screening system 10 described herein *vis*, the LEEP (Lactococcus Expression of Exported Proteins) system. These have been further analysed to remove artefacts. The nucleotide sequences of genes identified using the screening system have been characterised using a number of parameters described below.

15 1. All putative surface proteins are analysed for leader/signal peptide sequences. Bacterial signal peptide sequences share a common design. They are characterised by a short positively charged N-terminus (N region) immediately preceding a stretch of hydrophobic residues (central portion-h region) followed by a more polar C-terminal portion which contains the cleavage site (c-region). Computer 20 software is used to perform hydropathy profiling of putative proteins (Marcks, *Nuc. Acid. Res.*, 16:1829-1836 (1988)) which is used to identify the distinctive hydrophobic portion (h-region) typical of leader peptide sequences. In addition, the presence/absence of a potential ribosomal binding site (Shine-Dalgarno sequence required for translation) is also noted.

25 2. All putative surface protein sequences are used to search the OWL sequence database which includes a translation of the GENBANK and SWISSPROT database.. This allows identification of similar sequences which may have been previously characterised not only at the sequence level but at a functional level. It

may also provide information indicating that these proteins are indeed surface related and not artefacts.

3. Putative *S. agalactiae* surface proteins are also assessed for their novelty. Some of the identified proteins may or may not possess a typical leader peptide sequence and may not show homology with any DNA/protein sequences in the database. Indeed these proteins may indicate the primary advantage of our screening method, i.e. isolating atypical surface-related proteins, which would have been missed in all previously described screening protocols.

- 10 The construction of three reporter vectors and their use in *L. lactis* to identify and isolate genomic DNA fragments from pathogenic bacteria encoding secreted or surface associated proteins is now described.

Construction of the pTREP1-nuc series of reporter vectors

15 (a) **Construction of expression plasmid pTREP1**

The pTREP1 plasmid is a high-copy number (40-80 per cell) theta-replicating gram positive plasmid, which is a derivative of the pTREX plasmid which is itself a derivative of the previously published pIL253 plasmid. pIL253 incorporates the broad Gram-positive host range replicon of pAMβ1 (Simon and Chopin, *Biochemie* 20: 559-566 (1988))*L. lactis* sex-factor. pIL253 also lacks the *tra* function which is necessary for transfer or efficient mobilisation by conjugative parent plasmids exemplified by pIL501. The Enterococcal pAMβ1 replicon has previously been transferred to various species including *Streptococcus*, *Lactobacillus* and *Bacillus* 25 species as well as *Clostridium acetobutylicum*, (LeBlanc *et al.*, *Proceedings of the National Academy of Science USA* 75:3484-3487 (1978)) indicating the potential broad host range utility. The pTREP1 plasmid represents a constitutive transcription vector.

The pTREX vector was constructed as follows. An artificial DNA fragment containing a putative RNA stabilising sequence, a translation initiation region (TIR), a multiple cloning site for insertion of the target genes and a transcription terminator was created by annealing 2 complementary oligonucleotides and extending with Tfl DNA polymerase. The sense and anti-sense oligonucleotides contained the recognition sites for NheI and BamHI at their 5' ends respectively to facilitate cloning. This fragment was cloned between the XbaI and BamHI sites in pUC19NT7, a derivative of pUC19 which contains the T7 expression cassette from pLET1 (Wells *et al.*, *J. Appl. Bacteriol.* 74:629-636 (1993)) cloned between the EcoRI and HindIII sites. The resulting construct was designated pUCLEX. The complete expression cassette of pUCLEX was then removed by cutting with HindIII and blunting followed by cutting with EcoRI before cloning into EcoRI and SacI (blunted) sites of pIL253 to generate the vector pTREX (Wells and Schofield, *In Current advances in metabolism, genetics and applications-NATO ASI Series. H* 98:37-62. (1996)). The putative RNA stabilising sequence and TIR are derived from the *Escherichia coli* T7 bacteriophage sequence and modified at one nucleotide position to enhance the complementarity of the Shine Dalgarno (SD) motif to the ribosomal 16s RNA of *Lactococcus lactis* (Schofield *et al.* pers. coms. University of Cambridge Dept. Pathology.).

A *Lactococcus lactis* MG1363 chromosomal DNA fragment exhibiting promoter activity which was subsequently designated P7 was cloned between the EcoRI and BglII sites present in the expression cassette, creating pTREX7. This active promoter region had been previously isolated using the promoter probe vector pSB292 (Waterfield *et al.*, *Gene* 165:9-15 (1995)). The promoter fragment was amplified by PCR using the Vent DNA polymerase according to the manufacturer.

The pTREP1 vector was then constructed as follows. An artificial DNA fragment which included a transcription terminator, the forward pUC sequencing primer, a promoter multiple cloning site region and a universal translation stop sequence was created by annealing two overlapping partially complementary synthetic oligonucleotides together and extending with sequenase according to manufacturers instructions. The sense and anti-sense (pTREPF and pTREPR) oligonucleotides contained the recognition sites for EcoRV and BamHI at their 5' ends respectively to facilitate cloning into pTREX7. The transcription terminator was that of the *Bacillus penicillinase* gene, which has been shown to be effective in *Lactococcus* (Jos *et al.*, 5 *Applied and Environmental Microbiology* 50:540-542 (1985)). This was considered necessary as expression of target genes in the pTREX vectors was observed to be leaky and is thought to be the result of cryptic promoter activity in the origin region (Schofield *et al.* pers. coms. University of Cambridge Dept. Pathology.). The forward pUC primer sequencing was included to enable direct sequencing of cloned 10 DNA fragments. The translation stop sequence which encodes a stop codon in 3 different frames was included to prevent translational fusions between vector genes and cloned DNA fragments. The pTREX7 vector was first digested with EcoRI and blunted using the 5' - 3' polymerase activity of T4 DNA polymerase (NEB) according to manufacturer's instructions. The EcoRI digested and blunt ended 15 pTREX7 vector was then digested with Bgl II thus removing the P7 promoter. The artificial DNA fragment derived from the annealed synthetic oligonucleotides was then digested with EcoRV and Bam HI and cloned into the EcoRI(blunted)-Bgl II digested pTREX7 vector to generate pTREP. A *Lactococcus lactis* MG1363 chromosomal promoter designated P1 was then cloned between the EcoRI and BglII 20 sites present in the pTREP expression cassette forming pTREP1. This promoter was also isolated using the promoter probe vector pSB292 and characterised by Waterfield *et al.*, (1995) [*supra*]. The P1 promoter fragment was originally 25 amplified by PCR using vent DNA polymerase according to manufacturers

instructions and cloned into the pTREX as an EcoRI-BglII DNA fragment. The EcoRI-BglII P1 promoter containing fragment was removed from pTREX1 by restriction enzyme digestion and used for cloning into pTREP (Schofield *et al.* pers. coms. University of Cambridge, Dept. Pathology.).

5

(b) PCR amplification of the *S. aureus nuc* gene.

The nucleotide sequence of the *S. aureus nuc* gene (EMBL database accession number V01281) was used to design synthetic oligonucleotide primers for PCR amplification. The primers were designed to amplify the mature form of the *nuc* gene designated *nucA* which is generated by proteolytic cleavage of the N-terminal 19 to 21 amino acids of the secreted propeptide designated Snase B (Shortle, 1983 [*supra*]). Three sense primers (*nucS1*, *nucS2* and *nucS3*, shown in figure 3) were designed, each one having a blunt-ended restriction endonuclease cleavage site for EcoRV or SmaI in a different reading frame with respect to the *nuc* gene. Additionally BglII and BamHI were incorporated at the 5' ends of the sense and anti-sense primers respectively to facilitate cloning into BamHI and BglII cut pTREP1. The sequences of all the primers are given in figure 3. Three *nuc* gene DNA fragments encoding the mature form of the nuclease gene (*NucA*) were amplified by PCR using each of the sense primers combined with the anti-sense primer. The *nuc* gene fragments were amplified by PCR using *S. aureus* genomic DNA template, Vent DNA Polymerase (NEB) and the conditions recommended by the manufacturer. An initial denaturation step at 93°C for 2 min was followed by 30 cycles of denaturation at 93°C for 45 sec, annealing at 50°C for 45 seconds, and extension at 73°C for 1 minute and then a final 5 min extension step at 73°C. The PCR amplified products were purified using a Wizard clean up column (Promega) to remove unincorporated nucleotides and primers.

(c) Construction of the pTREP1-nuc vectors

The purified *nuc* gene fragments described in section b were digested with Bgl II and BamHI using standard conditions and ligated to BamHI and BglII cut and dephosphorylated pTREP1 to generate the pTREP1-*nuc*1, pTREP1-*nuc*2 and pTREP1-*nuc*3 series of reporter vectors. These vectors are described in figure 4. General molecular biology techniques were carried out using the reagents and buffers supplied by the manufacturer or using standard techniques (Sambrook and Maniatis, Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press: Cold Spring Harbour (1989)). In each of the pTREP1-*nuc* vectors the expression cassette comprises a transcription terminator, lactococcal promoter P1, unique cloning sites (Bgl II, EcoRV or SmaI) followed by the mature form of the *nuc* gene and a second transcription terminator. Note that the sequences required for translation and secretion of the *nuc* gene were deliberately excluded in this construction. Such elements can only be provided by appropriately digested foreign DNA fragments (representing the target bacterium) which can be cloned into the unique restriction sites present immediately upstream of the *nuc* gene.

(d) Screening for secreted proteins in Group B Streptococcus.

Genomic DNA isolated from Group B Streptococcus (*S. agalactiae*) was digested with the restriction enzyme Tru9I. This enzyme which recognises the sequence 5'-TTAA -3' was used because it cuts A/T rich genomes efficiently and can generate random genomic DNA fragments within the preferred size range (usually averaging 0.5 - 1.0 kb). This size range was preferred because there is an increased probability that the P1 promoter can be utilised to transcribe a novel gene sequence. However, the P1 promoter may not be necessary in all cases as it is possible that many Streptococcal promoters are recognised in *L. lactis*. DNA fragments of different size ranges were purified from partial Tru9I digests of *S. agalactiae* genomic DNA. As

the Tru 9I restriction enzyme generates staggered ends the DNA fragments had to be made blunt ended before ligation to the EcoRV or SmaI cut pTREP1-*nuc* vectors. This was achieved by the partial fill-in enzyme reaction using the 5'-3' polymerase activity of Klenow enzyme. Briefly Tru9I digested DNA was dissolved in a solution (usually between 10-20 μ l in total) supplemented with T4 DNA ligase buffer (New England Biolabs; NEB) (1X) and 33 μ M of each of the required dNTPs, in this case dATP and dTTP. Klenow enzyme was added (1 unit Klenow enzyme (NEB) per μ g of DNA) and the reaction incubated at 25°C for 15 minutes. The reaction was stopped by incubating the mix at 75°C for 20 minutes. EcoRV or SmaI digested pTREP-*nuc* plasmid DNA was then added (usually between 200-400 ng). The mix was then supplemented with 400 units of T4 DNA ligase (NEB) and T4 DNA ligase buffer (1X) and incubated overnight at 16°C. The ligation mix was precipitated directly in 100% Ethanol and 1/10 volume of 3M sodium acetate (pH 5.2) and used to transform *L. lactis* MG1363 (Gasson, *J. Bacteriol.* **154**:1-9 (1983)). Alternatively, the gene cloning site of the pTREP-*nuc* vectors also contains a BglII site which can be used to clone for example Sau3AI digested genomic DNA fragments.

L. lactis transformant colonies were grown on brain heart infusion agar and nuclease secreting (*Nuc*⁺) clones were detected by a toluidine blue-DNA-agar overlay (0.05 M Tris pH 9.0, 10 g of agar per litre, 10 g of NaCl per liter, 0.1 mM CaCl₂, 0.03 % wt/vol. salmon sperm DNA and 90 mg of Toluidine blue O dye) essentially as described by Shortle, 1983 [*supra*], and Le Loir *et al.*, 1994 [*supra*]). The plates were then incubated at 37°C for up to 2 hours. Nuclease secreting clones develop an easily identifiable pink halo. Plasmid DNA was isolated from *Nuc*⁺ recombinant *L. lactis* clones and DNA inserts were sequenced on one strand using the *NucSeq* sequencing primer described in figure 3, which sequences directly through the DNA insert.

Example 2**Preparation of a *S. agalactiae* standard inoculum****5 Strain validation**

S. agalactiae serotype III (strain 97/0099) is a recent clinical isolate derived from the cerebral spinal fluid of a new born baby suffering from meningitis. This haemolytic strain of Group B Streptococcus was epidemiologically tested and validated at the Respiratory and Systemic Infection Laboratory, PHLS Central Public Health

10 Laboratory, 61 Colindale Avenue, London NW9 5HT. The strain was subcultured only twice prior to its arrival in the laboratory. Upon its arrival on an agar slope, a sweep of 4-5 colonies was immediately used to inoculate a Todd Hewitt/5% horse blood broth which was incubated overnight statically at 37°C. 0.5 ml aliquots of this overnight culture were then used to make 20% glycerol stocks of the bacterium for 15 long-term storage at -70°C. Glycerol stocks were streaked on Todd Hewitt/5% horse blood agar plates to confirm viability.

***In vivo* passaging of Group B Streptococcus**

A frozen culture (described under strain validation) of *S. agalactiae* serotype III (strain 97/0099) was streaked to single colonies on Todd-Hewitt/5% blood agar plates, which were incubated overnight at 37°C. A sweep of 4-5 colonies was used to inoculate a Todd Hewitt/5% horse blood broth, which was again incubated overnight. A 0.5 ml aliquot from this overnight culture was used to inoculate a 50 ml Todd Hewitt broth (1:100 dilution) which was incubated at 37°C. 10-fold serial dilutions of the overnight culture were made (since virulence of this strain was unknown) and each was passaged intra-peritoneally (IP) in CBA/ca mice in duplicate. Viable counts were performed on the various inocula used in the passage. Groups of mice were challenged with various concentrations of the pathogen ranging from 10^8 to 10^4 colony forming units (cfu). Mice that developed symptoms were 30 terminally anaesthetized and cardiac punctures were performed (Only mice that had

been challenged with the highest doses, i.e. 1×10^8 cfu, developed symptoms). The retrieved unclotted blood was used to inoculate directly a 50ml serum broth (Todd Hewitt/20% inactivated foetal calf serum). The culture was constantly monitored and allowed to grow to late logarithmic phase. The presence of blood in the medium
5 interfered with OD_{600nm} readings as it was being increasingly lysed with increasing growth of the bacterium, hence the requirement to constantly monitor the culture. Upon reaching late logarithmic phase/early stationary phase, the culture was transferred to a fresh 50 ml tube in order to exclude dead bacterial cells and remaining blood cells which would have sedimented at the bottom of the tube. 0.5
10 ml aliquots were then transferred to sterile cryovials, frozen in liquid nitrogen and stored at -70°C. A viable count was carried out on a single standard inoculum aliquot in order to determine bacterial numbers. This was determined to be approximately 5 $X 10^8$ cfu per ml.

15 **Intra-peritoneal Challenge and virulence testing of Group B Streptococcus standard inoculum**

To determine if the standard inoculum was suitably virulent for use in a vaccine trial, challenges were carried out using a dose range. Frozen standard inoculum strain aliquots were allowed to thaw at room temperature. From viable count data the
20 number of cfu per ml was already known for the standard inoculum. Initially, serial dilutions of the standard inoculum were made in Todd Hewitt broth and mice were challenged intra-peritoneally with doses ranging from 1×10^8 to 1×10^4 cfu in a 500 µl volume of Todd Hewitt broth. The survival times of mouse groups injected with different doses of the bacterium were compared. The standard inoculum was
25 determined to be suitably virulent and a dose of 1×10^6 cfu was considered close to optimal for further use in vaccine trials. Further optimisation was carried out by comparing mice challenged with doses ranging between 5×10^5 and 5×10^6 cfu. The optimal dose was estimated to be approximately 2.5×10^6 cfu. This represented

a 100% lethal dose and was repeatedly consistent with end-points as determined by survival times being clustered within a narrow time-range. Throughout all these experiments, challenged mice were constantly monitored to clarify symptoms, stages of symptom development as well as calculating survival times.

5

Screening Group B Streptococcal LEEP derived genes in DNA vaccination experiments.

pcDNA3.1+ as a DNA vaccine vector

10 The commercially available pcDNA3.1+ plasmid (Invitrogen), referred to as pcDNA3.1 henceforth, was used as a vector in all DNA immunisation experiments involving gene targets derived using the LEEP system unless stated otherwise. pcDNA 3.1 is designed for high-level stable and transient expression in mammalian cells and has been used widely and successfully as a host vector to test candidate 15 genes from a variety of pathogens in DNA vaccination experiments (Zhang *et al.*, *Infection and Immunity* 176: 1035-40 (1997); Kurar and Splitter, *Vaccine* 15: 1851-57 (1997); Anderson *et al.*, *Infection and Immunity* 64: 3168-3173 (1996)).

20 The vector possesses a multiple cloning site which facilitates the cloning of multiple gene targets downstream of the human cytomegalovirus (CMV) immediate-early promoter/enhancer which permits efficient, high-level expression of the target gene in a wide variety of mammalian cells and cell types including both muscle and immune cells. This is important for optimal immune response as it remains unknown as to which cells types are most important in generating a protective response *in* 25 *vivo*. The plasmid also contains the ColE1 origin of replication which allows convenient high-copy number replication and growth in *E. coli* and the ampicillin resistance gene (B-lactamase) for selection in *E. coli*. In addition pcDNA 3.1

possesses a T7 promoter/priming site upstream of the MCS which allows for *in vitro* transcription of a cloned gene in the sense orientation.

Preparation of DNA vaccines

5 Oligonucleotide primers were designed for each individual gene of interest derived using the LEEP system unless stated otherwise. Each gene was examined thoroughly, and where possible, primers were designed such that they targeted that portion of the gene believed to encode only the mature portion of the protein (APPENDIX I); the intention being to express those sequences that encode only the
10 mature portion of a target gene protein to would facilitate its correct folding when expressed in mammalian cells. For example, in the majority of cases primers were designed such that putative N-terminal signal peptide sequences would not be included in the final amplification product to be cloned into the pcDNA3.1 expression vector. The signal peptide directs the polypeptide precursor to the cell
15 membrane via the protein export pathway where it is normally cleaved off by signal peptidase I (or signal peptidase II if a lipoprotein). Hence the signal peptide does not make up any part of the mature protein whether it be displayed on the bacterium's surface or secreted. Where an N-terminal leader peptide sequence was not immediately obvious, primers were designed to target the whole of the gene
20 sequence for cloning and ultimately, expression in pcDNA3.1.

All forward and reverse oligonucleotide primers incorporated appropriate restriction enzyme sites to facilitate cloning into the pcDNA3.1 MCS region. All forward primers were also designed to include the conserved Kozak nucleotide sequence 5'-
25 gccacc-3' immediately upstream of an 'atg' translation initiation codon in frame with the target gene insert. The Kozak sequence facilitates the recognition of initiator sequences by eukaryotic ribosomes. Typically, a forward primer incorporating a BamH1 restriction enzyme site the primer would begin with the sequence 5'-

cgggatccgccaccatg-3', followed by a sequence homologous to the 5' end of that part of a gene being amplified. All reverse primers incorporated a Not I restriction enzyme site sequence 5' -ttgcggccgc-3'. All gene-specific forward and reverse primers were designed with compatible melting temperatures to facilitate their amplification.

All gene targets were amplified by PCR from *S. agalactiae* genomic DNA template using Vent DNA polymerase (NEB) or rTth DNA polymerase (PE Applied Biosystems) using conditions recommended by the manufacturer. A typical amplification reaction involved an initial denaturation step at 95°C for 2 minutes followed by 35 cycles of denaturation at 95°C for 30 seconds, annealing at the appropriate melting temperature for 30 seconds, and extension at 72°C for 1 minute (1 minute per kilobase of DNA being amplified). This was followed by a final extension period at 72°C for 10 minutes. All PCR amplified products were extracted once with phenol chloroform (2:1:1) and once with chloroform (1:1) and ethanol precipitated. Specific DNA fragments were isolated from agarose gels using the QIAquick Gel Extraction Kit (Qiagen). The purified amplification gene DNA fragments were digested with the appropriate restriction enzymes and cloned into the pcDNA3.1 plasmid vector using *E. coli* as a host. Successful cloning and maintenance of genes was confirmed by restriction mapping and by DNA sequencing. Recombinant plasmid DNA was isolated on a large scale (> 1.5 mg) using Plasmid Mega Kits (Qiagen).

DNA vaccination trials

DNA vaccine trials in mice were accomplished by the administration of DNA to 6 week old CBA/ca mice (Harlan, UK). Mice to be vaccinated were divided into groups of six and each group was immunised with recombinant pcDNA3.1 plasmid DNA containing a specific target-gene sequence derived using the LEEP system unless stated otherwise. A total of 100 µg of DNA in Dulbecco's PBS (Sigma) was

injected intramuscularly into the tibialis anterior muscle of both hind legs. Four weeks later this procedure was repeated using the same amount of DNA. For comparison, control mice groups were included in all vaccine trials. These control groups were either not DNA-vaccinated or were immunised with non-recombinant pcDNA3.1 plasmid DNA only, using the same time course described above. Four weeks after the second immunisation, all mice groups were challenged intra-peritoneally with a lethal dose of *S. agalactiae* serotype III (strain 97/0099). The actual number of bacteria administered was determined by plating serial dilutions of the inoculum on Todd-Hewitt/5% blood agar plates. All mice were killed 3 or 4 days after infection. During the infection process, challenged mice were monitored for the development of symptoms associated with the onset of *S. agalactiae* induced-disease. Typical symptoms in an appropriate order included piloerection, an increasingly hunched posture, discharge from eyes, increased lethargy and reluctance to move which was often the result of apparent paralysis in the lower body/hind leg region.

The latter symptoms usually coincided with the development of a moribund state at which stage the mice were culled to prevent further suffering. These mice were deemed to be very close to death, and the time of culling was used to determine a survival time for statistical analysis. Where mice were found dead, a survival time was calculated by averaging the time when a particular mouse was last observed alive and the time when found dead, in order to determine a more accurate time of death. The results of this trial are shown in Table 1 and presented graphically in Figure 2.

Interpretation of Results

A positive result was taken as any DNA sequence that was cloned and used in challenge experiments as described above and gave protection against that challenge. DNA sequences were determined to be protective;

-if that DNA sequence gave statistically significant protection to mice as compared to control mice (to a 95% confidence level ($p > 0.05$) as determined using the Mann-Whitney U test .

5 -if that DNA sequence was marginal or non-significant using Mann-Whitney but showed some protective features. For example, one or more outlying mice may survive for significantly longer time periods when compared with control mice. Alternatively, the time to first death may also be prolonged when compared to counterpart mice in control groups. It is acceptable to allow marginal or non-significant results to be considered as potential positives when it is possible that the 10 clarity of some results may be affected by problems associated with the administration of the DNA vaccine. Indeed, much varied survival times may reflect different levels of immune response between different members of a given group.

Table 1

15 **LEEP DNA immunisation and GBS challenge Experiment****Statistical analysis of survival times**

	Mean Survival Times (hours)		
	UnVacc	3-60(ID-65)	3-5(ID-66)
1	27.583	54.416	42.916
2	27.583	31.000	42.916
3	24.583	43.000	32.874
4	22.250	34.916	42.916
5	35.916	38.958	27.333
6	22.250	34.916	30.916
Mean	27.583	40.458	37.791
sd	5.1691	8.9959	7.2860
p value		0.0098	0.0215

20 **p value** refers to statistical significance when compared to unvaccinated controls.

Comment**ID-65 (3-60)**

5

Mice immunised with the '3-60 (ID-65)' DNA vaccine exhibited significantly longer survival times when compared with the unvaccinated control group.

ID-66 (3-5)

10 Mice immunised with the '3-5 (ID-66)' DNA vaccine exhibited significantly longer survival times when compared with the unvaccinated control group.

Example 3

15 **Expression and Screening Group B Streptococcal LEEP derived Proteins in Protein vaccination experiments.**

Expression of proteins

Prioritised genes ie, those selected on the basis of predicted expression features as deduced from sequence characteristics (as described in Figure 1), were cloned and expressed as recombinant proteins using the pET system (Novagen, Inc., Madison, WI) utilising *Escherichia coli* as a host. Target genes were cloned into the pET28b(+) plasmid expression vector. The pET28b(+) vector is designed for high level expression and purification of target proteins. This vector carries a T7 promoter for transcription of a target gene, followed by an N-terminal His•Tag[®]/thrombin/T7•Tag[®] configuration, a multi-cloning site containing unique restriction enzyme sites for cloning purposes, and an optional C-terminal His•Tag sequence. The vector also carries a kanamycin resistance gene for selection purposes and for maintaining target gene expression (pET System Manual, 8th edition, Novagen).

Preparation of protein vaccines

Oligonucleotide primers were designed for each individual target gene derived using the LEEP system unless stated otherwise. Each gene was examined thoroughly. Where possible primers were designed so that they would target that part of the gene predicted to encode only the mature portion of the protein (APPENDIX II). It is hoped that expressing those corresponding to the predicted mature protein only, might facilitate its correct folding when finally expressed *in vitro*. Oligonucleotide primers were designed so that sequences, encoding the putative N-terminal signal peptide of the target protein, would not be included in the final amplification product to be cloned pET28b(+). The signal peptide directs the polypeptide precursor to the cell membrane via the protein export pathway where it is normally cleaved off by signal peptidase I (or signal peptidase II if a lipoprotein). Hence the signal peptide would not be expected to form any part of the mature target protein, whether it be displayed on the bacterium's surface or secreted. For this purpose, classical signal peptides and their cleavage sites were predicted using the DNA Strider™ Program (CEA, France) and the SignalP V1.1 program, which predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms (Nielsen *et al.*, *Protein Engineering* 10: 1-6 (1997)). Where a N-terminal leader peptide sequence was not obvious, primers were designed to include the whole of the gene sequence for cloning and expression.

All oligonucleotide primers were designed to incorporate appropriate restriction enzyme sites to facilitate cloning into the pcDNA3.1 MCS region (APPENDIX II). Forward primers included an *Nco* I (5'-ccatgg-3') or *Nhe* I (5'-gcttagc-3') restriction enzyme site and an 'ATG' start codon in-frame with the target gene open reading frame (orf). All reverse primers included a *Not* I restriction enzyme site 5' - gccccgc-3' and were designed so that the target gene could be expressed in frame with the C-terminal His•Tag (i.e. the stop codon of the target gene was not

included). Using the *Nco* I and *Not* I, allowed the removal of the N-terminal His•Tag[®], thrombin and T7•Tag[®] DNA sequences. At the same time target genes were cloned immediately downstream of a highly efficient ribosome binding site (from the phage T7 major capsid protein), to facilitate high level expression/translation of the target gene by T7 RNA polymerase, and subsequent purification by means of the C-terminal His•Tag. All target gene-specific forward and reverse primers were designed with compatible melting temperatures to facilitate their amplification.

All gene targets were amplified by PCR from *S. agalactiae* genomic DNA template using Vent DNA polymerase (NEB) using conditions recommended by the manufacturer. A typical amplification reaction involved an initial denaturation step at 95°C for 2 minutes followed by 35 cycles of denaturation at 95°C for 30 seconds, annealing at the appropriate melting temperature for 30 seconds, and extension at 72°C for 1 minute (1 minute per kilobase of DNA being amplified). This was followed by a final extension period at 72°C for 10 minutes. All PCR amplified products were extracted once with phenol:chloroform (2:1:1) and once with chloroform (1:1) and ethanol precipitated. Specific DNA fragments were isolated from agarose gels using the QIAquick Gel Extraction Kit (Qiagen). Purified target gene DNA amplicons were then digested *Nco* I (or *Nhe* I) and *Not* I restriction enzymes, and cloned into *Nco* I and *Not* I digested pET28b(+) plasmid vector using *E. coli* DH5 α or *E. coli* BL21 (DE3) as a host. Successful cloning and maintenance of genes was confirmed by restriction mapping.

Determination of target protein expression and solubility

Glycerol stocks of *E. coli* BL21 DE3 pET28b(+) strains expressing recombinant proteins were used to inoculate 10 ml Luria broth containing Kanamycin (30 μ g/ml) which were grown overnight at 37°C with vigorous shaking (300 rpm).

A 20-40 ml Luria broth containing Kanamycin (30 µg/ml) was inoculated with 1:100 dilution of the overnight culture from step 1 and grown at 37°C with vigorous shaking (300 rpm). When the culture reached an OD₆₀₀ of between 0.6 and 1.0, IPTG was added to a final concentration of 1mM. Typically cultures were induced 5 for 3 hours. Cells were then harvested by centrifugation at 7000 g for 10 min. The cell pellet was then resuspended in 1/10 volume of lysis buffer (50mM NaH₂PO₄, pH.8.0; 300mM NaCl;10mM imidazole; 10% glycerol). Lysozyme was then added to a final concentration of 1mg/ml, and the suspension was incubated on ice for 30 min. The suspension was then sonicated on ice (six 10-sec bursts at 200-300 W with 10 a 10-sec cooling period. The lysate was then centrifuged at 10,000g for 20 min. The supernatant (containing soluble protein) was transferred to a sterile 2 ml eppendorf. The pellet was resuspended in 2 ml of solubilisation buffer (8 M Urea; 50mM NaH₂PO₄, pH.8.0; 300mM NaCl; 10% glycerol). This suspension contained the 15 insoluble protein fraction. Aliquots from both the soluble and insoluble fractions were transferred to new eppendorfs. The protein samples were denatured by adding an equal volume of 2x SDS-PAGE buffer and heating at 95°C for 5 min. Denatured extract samples were then analysed by SDS-PAGE to determine target gene expression and solubility.

Large scale expression of recombinant target proteins

Glycerol stocks of *E. coli* BL21 DE3 pet28b(+) strains expressing recombinant 5 proteins were used to inoculate 10 ml Luria broth containing Kanamycin (30 µg/ml) which were grown overnight at 37°C with vigorous shaking (300 rpm). 5 ml of an overnight culture of a recombinant strain was used to inoculate a 250 ml Luria broth containing kanamycin (30 µg/ml) which was grown at 37°C with vigorous shaking (300 rpm). When the culture reached an OD₆₀₀ of between 0.6 and 1.0, IPTG was 10 added to a final concentration of 1mM. Typically, cultures were induced for 3 hours. Cultures were then centrifuged to a pellet and stored frozen at -20°C.

Purification of target antigens.

15 Ni-NTA agarose (Qiagen LTD, West Sussex, UK; Cat. No. 30210) was used to purify the His-Tagged recombinant proteins. The 6xHis affinity tag which was expressed in frame with the target proteins in pET28b(+), facilitates binding to Ni-NTA. Ni-NTA offers high binding capacity (with minimal non-specific binding) and can bind 5-10 mg of 6xHis-tagged protein per ml of resin. The 6xHis-tag is poorly 20 immunogenic, and at pH 8.0, the tag is small, uncharged and therefore does not generally interfere with the structure and function of the protein (The QIAexpressionist, Qiagen Handbook, March 1999).

25 NOTE: All the proteins (LEEP-derived, unless stated otherwise) described here were purified under denaturing conditions except ID-65. ID-65 was prepared and purified under native conditions.

Purification under native conditions

30 The frozen pellet was allowed to thaw on ice for 15 minutes and then resuspended in 10 ml of lysis buffer (50mM NaH₂PO₄, pH.8.0; 300mM NaCl;10mM imidazole;

10% glycerol). Lysozyme was then added to a final concentration of 1mg/ml, and the suspension was incubated on ice for 30 min. The suspension was then sonicated on ice (six 10-sec bursts at 200-300 W with a 10-sec cooling period). Dnase I (5 µg/ml) was then added to the lysate, which was then incubated on ice for 10-15 min.

5 The lysate was then centrifuged at 10,000 rpm for 20 min at 4°C to pellet cell debris. The clear lysate supernatant was then loaded into a polypropylene column (Qiagen; Cat. No. 34964), bottom cap attached. 1.5 ml of 50% Ni-NTA was then added, the column sealed and the suspension was allowed to mix gently using a rotating wheel for 1-2 hours at 4°C. The column containing the lysate/Ni-NTA mix was then

10 placed upright using a retort stand, and the Ni-NTA was allowed to settle. The bottom cap was removed and the lysate was allowed to flow through. The column was then washed with three to six 4 ml volumes of wash buffer (50mM NaH₂PO₄, pH.8.0; 300mM NaCl;20mM imidazole; 10% glycerol). The protein was then eluted in 0.5 ml aliquots of elution buffer (50mM NaH₂PO₄, pH.8.0; 300mM

15 NaCl;500mM imidazole; 10% glycerol). Eluate fractions were then analysed by SDS-PAGE and those containing the protein were pooled and dialysed against a PBS (pH 7.0)-glycerol (10%) solution.

Purification and refolding under denaturing conditions

20 The frozen pellet was allowed to thaw on ice for 15 minutes and then resuspended in 10 ml of buffer containing 8 M Urea, 300 mM NaCl, 10% glycerol, 0.1 M NaH₂PO₄, pH.8.0, and 10 mM imidazole. The cells were then lysed by gentle vortexing for 1 hour at room temperature. The lysate was then centrifuged at

25 10,000g for 20 minutes to pellet cellular debris. The clear lysate supernatant was then loaded into a polypropylene column (Qiagen; Cat. No. 34964), bottom cap attached. 1.5 ml of 50% Ni-NTA slurry was then added, the column sealed and the suspension was allowed to mix gently using a rotating wheel for 1-2 hours at room

- temperature. The column containing the lysate/Ni-NTA mix was then placed upright using a retort stand, and the Ni-NTA was allowed to settle. The bottom cap was removed and the lysate was allowed to flow through. The column was then washed with 4-8 ml of buffer containing 8 M Urea, 300 mM NaCl, 10% glycerol, 0.1 M NaH₂PO₄, pH 8.0, and 10 mM imidazole. The resin was then washed with a gradient of 6 to 0 M in a buffer containing 0.1 M NaH₂PO₄, pH 8.0, 300 mM NaCl and 10% glycerol to facilitate the slow removal of urea and gradual refolding of target protein. The resin was then washed with a buffer containing 0.1 M NaH₂PO₄, pH 7.0, 500 mM NaCl and 10% glycerol. The recombinant protein was then eluted 5 in 0.5 ml aliquots with 500 mM Imidazole in 0.1 mM NaH₂PO₄, pH 7.0, 500 mM NaCl and 10% glycerol. The fractions were analysed on SDS-PAGE and those containing the protein were pooled and dialysed against a PBS (pH 7.0)-glycerol 10 (10%) solution.
- 15 All purified proteins were analysed by SDS-PAGE, as shown in Figures 5, 6 and 7, prior to their use as antigens in immunisation and vaccination experiments.
- Protein Vaccinations**
- 20 Vaccines were composed of the target protein in phosphate buffered saline/10% glycerol and mixed with aluminium hydroxide (alum) (Imject[®] Alum, Pierce, Rockford, Ill.). Each dose (unless otherwise stated) of vaccine contained 25 µg of purified protein in 50 µl of PBS/10% glycerol, mixed with 50 µl of alum. Groups of 25 6-8 CBA/ca mice (Harlan, UK) were immunised subcutaneously with the vaccines and again 4 weeks later. A control group received 100 µl dose of PBS/10% glycerol with alum. All vaccinated groups consisted of 6 mice. Mice were challenged at 7 weeks (unless otherwise stated). Mice were injected intraperitoneally (i.p.) with between 2.5-5 X 10⁶ bacteria diluted in 0.5 ml Todd-Hewitt broth. Deaths were recorded daily for 7 days. The challenged mice were observed daily for signs of 30 illness. Typical symptoms in an appropriate order included piloerection, an

increasingly hunched posture, discharge from eyes, increased lethargy and reluctance to move which was often the result of apparent paralysis in the lower body/hind leg region. The latter symptoms usually coincided with the development of a moribund state at which stage the mice were culled to prevent further suffering. These mice
5 were deemed to be very close to death, and the time of culling was used to determine a survival time for statistical analysis. Where mice were found dead, a survival time was calculated by averaging the time when a particular mouse was last observed alive and the time when found dead, in order to determine a more accurate time of death.

10

Analysis of antibody responses

Mice (6 per group) were immunised with two doses of vaccine with a four week interval. Mice were tail bled at 3 weeks and 6 weeks post primary vaccination to
15 obtain sera. Total Immunoglobulin G (IgG) titres to the vaccine protein component in sera were determined by enzyme-linked immunosorbent assay (ELISA), using the original purified protein as the coating antigen.

Standard ELISA protocol

20

Solutions

Carbonate/bicarbonate buffer, pH 9.8

0.80g Na₂CO₃

1.46g NaHCO₃

pH to 9.6 using HCl

25

Add distilled water (dH₂O) to a final volume of 500ml.

n-NITROPHENYL PHOSPHATE SUBSTRATE

Diethanolamine Buffer, pH 9.8

48.5 ml diethanolamine

pH to 9.8 using 1M HCl

Add dH₂O to a final volume of 500ml

5

NOTE: ELISAs were optimised for each protein submitted for immunisation.

PROTOCOL

1. ELISA plates (Greiner labortechnik 96 well plates: Cat. No. 655061) with an appropriate concentration of recombinant protein diluted in carbonate/bicarbonate buffer (50 µl/well). Cover plates with plastic or foil and leave overnight at 4°C.
- 10 2. Quickly wash plates twice in a tub/container containing PBS/0.05% Tween-20 and then pat dry.
3. Block plates with 3% BSA in PBS/Tween (100µl /well) for 1 hour at room temperature.
- 15 4. Wash the plates 3 times PBS/Tween as before and pat dry as before.
5. Apply (primary antibody) protein-specific antiserum (50µl/well) diluted from 1/50 in a doubling dilution series in PBS/Tween and incubate at room temperature for 90 minutes.
- 20 6. Wash plates as before (3 times quickly), followed up by 2 X 3 minute soaks (in PBS/Tween)
7. Apply diluted secondary antibody alkaline phosphatase conjugate. For anti-mouse Total IgG alkaline phosphatase conjugate (Goat Anti-Mouse IgG-AP, Southern Biotechnology Associates, Birmingham, AL. Cat. No. 1030-04) dilute 1/3000 in PBS/Tween and apply 50 µl per well and incubate at room temperature for 90 minutes.
- 25 8. Wash plates as in step 6.

9. Apply substrate. Dissolve one 5mg tablet of nitrophenyl phosphate (Sigma:kept in freezer) in 5ml of diethanolamine buffer. Apply 100 μ l per well. Cover with foil (a light-sensitive reaction) and leave at room temperature for 30 minutes. Read Optical densities (OD) at a wavelength of 405nm.
- 5 10. Plot curves of OD Vs dilution (log scale). Calculate end-point titres as the dilution giving the same OD as the mean of the OD obtained from the wells containing the 1/50 dilution of pre-immune serum.

10

15

ELISA Plate format

2°	1/50	1/10	1/20	1/40	1/80	1/160	1/32	1/640	1/1280	1/256	1/5120
1°	Duplicate										
Pre											
Pre											
Pre											

Table Summary

5 **Pre** Replicate wells of pooled pre-inoculation serum (50µl per well) diluted to 1/50 are included on every plate in order for end point titres to be calculated.

2° Is a blank control well to which no secondary antibody conjugate is applied.

PBS/Tween by itself is applied instead

1° Is a blank control well to which no primary antibody is applied. PBS/Tween by itself is applied instead

Duplicate Each serum is analysed in duplicate

The dilution series used is indicated (see first row of table). Beginning with a 1/50 dilution, sera are diluted two-fold in PBS/Tween in doubling dilution series as indicated.

15

Protein Immunisation data**ID-65 and ID-83**

The ID-65 and ID-83 vaccines were composed of the target proteins in phosphate buffered saline/10% glycerol mixed with aluminium hydroxide (alum)

(Imject[®] Alum, Pierce, Rockford, Ill.). Each dose of vaccine contained 20 µg of purified protein in 100 µl of PBS/10% glycerol, mixed with 50 µl of alum. A group of 6-8 week old CBA/ca mice (Harlan, UK) were immunised subcutaneously with the ID-65 and ID-83 vaccine and again 4 weeks later. A control group received a 5 150 µl dose of PBS/10% glycerol (2:1) with alum. All groups consisted of 6 mice. Mice were tail bled at 5 weeks post primary vaccination to obtain sera. The presence of total Immunoglobulin G (IgG) antibodies to the ID-65 and ID-83 protein in sera was determined by enzyme-linked immunosorbent assay (ELISA), using the purified protein as the coating antigen. ELISA was also performed using sera obtained at 6 10 weeks post-primary vaccination from the PBS/10% glycerol immunised control group.

NOTE: ELISA plates were coated with the ID-65 or ID-83 proteins at a concentration of 1 µg/ml.

15

Protein Vaccination -ELISA results for ID-65 and ID-83

Mice (6 per group) were immunised with two doses of the ID-65 and ID-83 vaccines with a four week interval. Mice were tail bled at 5 weeks post primary 20 vaccination to obtain sera. The Immunoglobulin G (IgG) titres to the vaccine protein component in sera were determined by enzyme-linked immunosorbent assay (ELISA), using the purified ID-65 and ID-83 proteins as the coating antigen. Subsequent to optimisation, ELISA plates were coated at a concentration 1µg/ml for both the purified ID-65 and ID-93 proteins. Total IgG titres were measured against 25 pre-immune serum (1/50 dilution). The results are shown in Table 2 and graphically in Figure 8.

30

Table 2

Serum (Group)	<u>ID-65 + Alum</u> (n=6)	<u>PBS + Alum</u> (n=6)	<u>ID-83 + Alum</u> (n=6)	<u>PBS + Alum</u> (n=6)
Coating antigen	<u>ID-65</u>		<u>ID-83</u>	
Bleed	<u>5 weeks</u>	<u>5 weeks</u>	<u>5 weeks</u>	<u>5 weeks</u>
<u>Total IgG Titres</u> <u>(mouse 1-6)</u>	7535763	965	82081	61
	1557649	90	50027	50
	3319737	108	154670	80
	1832259	176	57901	96
	8794360	371	66497	125
	1445728	0	49928	0
<u>Average</u>	4080916	285	76851	69
<u>Standard Deviation</u>	3258818	355	39985	43

5

Protein Immunisation and Challenge data (ID-93)**ID-93**

The ID-93 vaccine was composed of the target ID-93 protein in phosphate buffered saline/10% glycerol mixed with aluminium hydroxide (alum) (Imject® Alum, Pierce, Rockford, Ill.). Each dose of vaccine contained 25 µg of purified protein in 100 µl of PBS/10% glycerol, mixed with 100 µl of alum. A group of 6-8 week old CBA/ca mice (Harlan, UK) were immunised subcutaneously with the ID-93 vaccine and again 4 weeks later. A control group received PBS/10% glycerol with alum. Both groups consisted of 6 mice. Mice were challenged at 7 weeks (unless otherwise stated). Mice were injected intraperitoneally (i.p.) with 5×10^6 bacteria diluted in

0.5 ml Todd-Hewitt broth. The challenged mice were observed daily for signs of illness. Deaths were recorded daily for 7 days. Survival data are shown in Table 3 and graphically in Figure 9.

- 5 Mice were tail bled at 3 weeks and 6 weeks post primary vaccination to obtain sera. The presence of total Immunoglobulin G (IgG) antibodies to the ID-93 protein in sera was determined by enzyme-linked immunosorbent assay (ELISA), using the pure ID-93 protein as the coating antigen. ELISA was also performed using sera obtained at 6 weeks post-primary vaccination from the PBS/10% glycerol immunised
10 control group.

Note: ELISA plates were coated with the ID-93 protein at a concentration of 1 $\mu\text{g}/\text{ml}$.

15 **Table 3**
ID-93 protein immunisation and GBS challenge experiment

Statistical analysis of Survival Times

Group	PBS + Alum	ID-93 + Alum
<u>Survival Times (hours)</u>	22.37	29.37
	22.37	35.12
	15.37	32.62
	28.03	32.62
	29.53	37.12
	26.53	27.87
Mean	24.03	32.45
sd	5.16	3.45
p value		0.01

20

p value refers to statistical significance when compared to unvaccinated controls.

Comment**5 ID-93 (RS-70)**

Mice immunised with the ID-93-Alum vaccine exhibited significantly longer survival times when compared with the PBS-Alum control group.

(Statistical Significance was determined by the Mann-Whitney U test using a 95%
10 confidence level ($p > 0.05$).

Protein Vaccination -ELISA results for ID-93

Mice (6 per group) were immunised with two doses of the ID-93 vaccine with a four
15 week interval. Mice were tail bled at 3 weeks and 6 weeks post primary vaccination
to obtain sera. The Immunoglobulin G (IgG) titres to the vaccine protein component
in sera were determined by enzyme-linked immunosorbent assay (ELISA), using the
purified ID-93 protein as the coating antigen. Subsequent to optimisation, ELISA
plates were coated with the purified ID-93 protein at a concentration of 1 μ g/ml.
20 Total IgG titres were measured against pre-immune serum (1/50 dilution). The
results are shown in Table 4 and graphically in Figure 10.

Table 4

Serum Group	<u>ID-93 + Alum(n=6)</u>		<u>PBS/10%glycerol (n=6)</u> (control)	
Coating antigen	<u>ID-93</u>	<u>ID-93</u>	<u>ID-93</u>	<u>ID-93</u>
Bleed	<u>3 weeks</u>	<u>6 weeks</u>	<u>3 weeks</u>	<u>6 weeks</u>
<u>Total IgG Titres (mouse 1-6)</u>	87196	3000000	39	100
	99544	8000000	31	16
	19620	2000000	31	79
	34724	10000000	59	48
	59990	10000000	24	328
	30041	4000000	13	40
<u>Average</u>	55186	6166667	33	102
<u>Standard error</u>	32654	3600926	15	115

5

**Protein Immunisation data
ID-89 and ID-96**

The ID-89 and ID-96 vaccines were composed of the target proteins in phosphate buffered saline/10% glycerol mixed with TitreMax Gold adjuvant (Sigma, Missouri, USA) according to the manufacturers instructions. The ID-89 vaccine contained 25 µg of purified protein 50 µl of PBS/10% glycerol, mixed with 50 µl of TitreMax Gold. The ID-96 vaccine contained 12.5 µg of purified protein 50 µl of PBS/10% glycerol, mixed with 50 µl of TitreMax Gold. Groups of 6-8 week old CBA/ca mice (Harlan, UK) were immunised subcutaneously with the ID-89 and ID-96 vaccines and again 4 weeks later. A control group received a 100 µl dose PBS/10% glycerol with TitreMax Gold (1:1). Both groups consisted of 6 mice. Mice were tail bled at 3

weeks and 6 weeks post primary vaccination to obtain sera. The presence of total Immunoglobulin G (IgG) antibodies to the ID-65 and ID-83 protein in sera was determined by enzyme-linked immunosorbent assay (ELISA), using the purified protein as the coating antigen. ELISA was also performed using sera obtained at 3
5 weeks and 6 weeks post-primary vaccination from the PBS/10% glycerol immunised control group.

Note: ELISA plates were coated with the ID-89 or ID-96 proteins at a concentration of 1 μ g/ml and 3 μ g/ml respectively.

10

Protein Vaccination -ELISA results for ID-89 and ID-96

Mice (6 per group) were immunised with two doses of the ID-89 and ID-96 vaccines with a four week interval. Mice were tail bled at 3 weeks and 6 weeks post primary
15 vaccination to obtain sera. The Immunoglobulin G (IgG) titres to the vaccine protein component in sera were determined by enzyme-linked immunosorbent assay (ELISA), using the purified ID-65 and ID-83 proteins as the coating antigen. Subsequent to optimisation, ELISA plates were coated with purified ID-89 and ID-96 protein at a concentration 1 μ g/ml and 3 μ g/ml respectively. Total IgG titres were
20 measured against pre-immune serum (1/50 dilution). ELISA was also performed on both proteins using sera obtained at 3 weeks and 6 weeks post-primary vaccination from the PBS/10% glycerol immunised control group. Results are shown in tables 5a and 5b and graphically in Figure 11.

Table 5a

Serum	ID-89 + TitreMax Gold (n=6)		ID-96 + TitreMax Gold(n=6)	
Coating antigen	ID-89		ID-96	
Bleed	<u>3 weeks</u>	<u>6 weeks</u>	<u>3 weeks</u>	<u>6 weeks</u>
<u>Total IgG Titres (mouse 1- 6)</u>	146940	1000000	190371	10000000
	89672	1000000	212505	10000000
	173532	2000000	167613	5000000
	85161	751210	110378	5000000
	88956	551281	142614	1000000
	27880	2000000	191085	1000000
Average	102024	1217082	169094	5333333
Standard Deviation	51451	629364	37341	4033196

Table 5b

Serum	<u>PBS/10%glycerol (n=6)</u>		<u>PBS/10%glycerol (n=6)</u>	
Coating protein	<u>ID-89</u>		<u>ID-96</u>	
Bleed	<u>3 weeks</u>	<u>6 weeks</u>	<u>3 weeks</u>	<u>6 weeks</u>
<u>Total IgG Titres (mouse 1-6)</u>	3	7	33	31
	8	18	77	62
	29	31	77	1
	34	4	52	29
	0	2	125	31
	5	1	113	0
Average	13	11	80	26
Standard deviation	15	12	35	23

Example 4**Conservation and variability of candidate vaccine antigen genes among different isolates of Group B Streptococci**

An initial Southern blot analysis was carried out to determine cross-serotype conservation of novel Group B Streptococcal genes isolated using the LEEP system unless stated otherwise. Analysing the serotype distribution of a target gene will also determine their potential use as antigen components in a GBS vaccine. The Group B Streptococcal strains whose DNA was analysed as part of this study are listed in APPENDIX III

Amplification and labelling of specific target genes as DNA probes for Southern blot analysis.

Oligonucleotide primers were designed for each individual gene of interest derived using the LEEP system unless stated otherwise. The same primers already described in APPENDIX II were used to amplify corresponding gene-specific DNA probes. Specific gene targets were amplified by PCR using Vent DNA polymerase (NEB) according to the manufacturers instructions. Typical reactions were carried out in a 100 µl volume containing 50 ng of GBS template DNA, a one tenth volume of enzyme reaction buffer, 1 µM of each primer, 250 µM of each dNTP and 2 units of Vent DNA polymerase. A typical reaction contained an initial 2 minute denaturation at 95°C, followed by 35 cycles of denaturation at 95°C for 30 seconds, annealing at the appropriate melting temperature for 30 seconds, and extension at 72°C for 1 minute (1 minute per kilobase of DNA being amplified). The annealing temperature was determined by the lower melting temperature of the two oligonucleotide primers. The reaction was concluded with a final extension period of 10 minutes at 72°C.

All PCR amplified products were extracted once with phenol chloroform (2:1:1) and once with chloroform (1:1) and ethanol precipitated. Specific DNA fragments were isolated from agarose gels using the QIAquick Gel Extraction Kit (Qiagen). For use as DNA probes, purified amplified gene DNA fragments were labelled with digoxigenin using the DIG Nucleic Acid Labelling Kit (Boehringer Mannheim) according to the manufacturer's instructions.

Southern blot hybridisation analysis of Group B Streptococcal genomic DNA

Genomic DNA had previously been isolated from all strains of Group B Streptococci which were investigated for conservation of LEEP-derived (unless stated otherwise) gene targets. Appropriate DNA concentrations were digested using either *Hin* DIII or *Eco* RI restriction enzymes (NEB) according to manufacturer instructions and analysed by agarose gel electrophoresis. Following agarose gel electrophoresis of DNA samples, the gel was denatured in 0.25M HCl for 20 minutes and DNA was transferred onto HybondTM N⁺ membrane (Amersham) by overnight capillary blotting. The method is essentially as described in Sambrook *et al.* (1989) using Whatman 3MM wicks on a platform over a reservoir of 0.4M NaOH. After transfer, the filter was washed briefly in 2x SSC and stored at 4°C in Saran wrap (Dow chemical company).

Filters were prehybridised, hybridised with the digoxigenin labelled DNA probes and washed using conditions recommended by Boehringer Mannheim when using their DIG Nucleic Acid Detection Kit. Filters were prehybridised at 68°C for one hour in hybridisation buffer (1% w/v supplied blocking reagent, 5x SSC, 0.1% v/v N-lauryl sarcosine, 0.02% v/v sodium dodecyl sulphate[SDS]). The digoxigenin labelled DNA probe was denatured at 99.9°C for 10 minutes before being added to the hybridisation buffer. Hybridisation was allowed to proceed overnight in a rotating Hybaid tube in a Hybaid Mini-hybridisation oven. Unbound probe was removed by washing the filter twice with 2x SSC- 0.1% SDS for 5 minutes at room

temperature. For increased stringency filters were then washed twice with 0.1x SSC-0.1% SDS for 15 minutes at 68°C. The DIG Nucleic Acid Detection Kit (Boehringer Mannheim) was used to immunologically detect specifically bound digoxigenin labelled DNA probes.

5

Results of Southern blot analysis

Unless otherwise stated, all genomic digests and their corresponding Southern blots followed an identical lane order as described in Table 6 below.

10

Table 6

Lane	1	2	3	4	5	6	7
Strain	1 kb molecular	S15	A909	SB35	H36B	18RS21	1954/92
Serotype	Weight	Ia	Ia	Ib	Ib	II	II
	Marker						

Lane	8	9	10	11	12	13	14
Strain	118/158	97/0057	BS30	M781	97/0099	3139	1169-NT
Serotype	II	II	III	III	III	IV	V

Lane	15	16	17	18	19	20
Strain	GBS 6	7271	JM9	Group A Streptococcus	<i>Streptococcus</i> <i>pneumoniae</i>	1 kb molecular
Serotype	VI	VII	VIII	-	14	Weight Marker

15

For comparative purposes, it was decided to analyse the serotype distribution of the GBS *rib* gene, which encodes the known protective immunogen Rib. Rib has previously been shown to be present in serotype III and some strains of serotype II but not in serotypes Ia or Ib (Stalhammar-Carlemalm *et al.*, *J. Exp. Med.* 177: 1593-1603 (1993)).

Confirmation of this pattern would not only give increased confidence in interpreting subsequent results, it would also determine if a *rib* gene homologue was present in the remaining GBS serotypes being investigated here. Primers designed for the amplification of *rib* for use as a gene probe in Southern blot analysis are described in APPENDIX II.

Table 7 – Lane order for Figure 12 (*rib* gene Southern blot analysis)

Lane	1	2	3	4	5	6	7
Serotype	1 kb molecular marker	515	A909	SB35	H36B	18RS21	1954/92
Serotype	Weight Marker	Ia	Ia	Ib	Ib	II	II

15

Lane	8	9	10	11	12	13	14
Serotype	118/158	97/0057	BM110	BS30	M781	97/0099	3139
Serotype	II	II	III	III	III	III	IV

Lane	15	16	17	18	19	20
Serotype	1169-NT	GBS 6	7271	JM9	Group A Streptococcus pneumoniae	
Serotype	V	VI	VII	VIII	—	14

Rib (Figure 12) Comment

The Southern blot analysis shown in Figure 12 indicates that the *rib* gene is not conserved across all GBS serotypes. *rib* appears to be absent from all serotype Ia and Ib strains (lanes 2 to 5) and from strains 118/158 and 97/0057 of serotype II (lanes 8 and 9). However, *rib* would appear to present in strains 18RS21 and 1954/92 of serotype II (lanes 6 and 7) and in all strains of serotype III (lanes 10 to 13). This is in agreement with previously published data (Stalhammar-Carlemalm *et al.*, 1993 [*supra*]). *rib* would also appear to be present in strains representing serotypes VII and VIII (lanes 17 and 18) but was absent from strains representing serotypes IV, V and VI (lanes 14 to 16) as well as the control strains (lanes 19 and 20). The *rib* gene probe did hybridise with lower intensity to genomic DNA fragments from strains representing serotypes Ia, Ib, IV, VI, VII and serotype II strains 118/158 and 97/0057. This may indicate the presence of a gene in these strains with a lower level of homology to *rib*. These hybridising DNA fragments may contain a homologue of the GBS *bca* gene encoding the Ca protein antigen which has been shown to be closely homologous to the Rib protein (Wastfelt *et al.*, *J. Biol. Chem.* 271:18892-18897 (1996)). If this is the case, it would be in agreement with previous work which showed all strains of serotypes Ia, Ib, II and III to be positive for one of the two proteins (Stalhammar-Carlemalm *et al.*, 1993 [*supra*]). However, the apparent variable distribution of the *rib* gene amongst different GBS

serotypes, makes it a less than ideal candidate for use in a GBS vaccine that is cross-protective against all serotypes.

5

ID-65 (Figure 13) Comment

The Southern blot analysis described in Figure 13 indicates that gene ID-65 is conserved across all GBS serotypes. The gene probe hybridised specifically to a *Hin* DIII-digested genomic DNA fragment of approximately 3.0 kb in DNA digests from 10 all GBS representatives, and was absent from both the control strains (lanes 18 and 19). This would suggest that the ID-65 gene is conserved across all GBS serotypes (and strains) at both the gene and locus level. The ID-65 DNA probe also hybridised weakly to the 1.636 bp molecular weight marker (the 1 kb DNA ladder from NEB was used to estimate DNA fragment sizes in all Southern blot analyses).

15

ID-89 (Figure 14) Comment

The Southern blot analysis described in Figure 14 indicates that gene ID-89 may not be conserved across all GBS serotypes. A 4.0 kb *Hin*DIII-digested genomic DNA fragment from 12 out of 16 GBS strains hybridised specifically to the ID-89 gene probe. In addition, a 3.25 kb *Hin*DIII-digested genomic DNA fragment from the GBS strain Ib (SB35) [lane 4] also hybridised specifically with the ID-89 gene probe. However, the ID-89 gene probe did not hybridise to digested genomic DNA fragments from strains Ia (515) [lane 2], IV (3139) [lane 13] and V (1169-NT) [lane 14], suggesting that these strains do not possess a ID-89 gene homologue.

25

ID-93 (Figure 15) Comment

The Southern blot analysis described in Figure 15 indicates that gene ID-93 is conserved across all GBS serotypes. The gene probe hybridised specifically to a *Hin*

DIII-digested genomic DNA fragment of approximately 3.25 kb in DNA digests from all GBS representatives, and was absent from both the control strains (lanes 18 and 19). This would suggest that the ID-93 gene is conserved across all GBS serotypes (and strains) at both the gene and locus level.

5

ID-96 (Figure 16) Comment

The Southern blot analysis described in Figure 16 indicates that gene ID-96 is conserved across all GBS serotypes. The gene probe hybridised specifically to a *Eco* RI-digested genomic DNA fragment of approximately 12.0 kb in DNA digests from all GBS representatives, and was absent from both the control strains (lanes 18 and 19). This would suggest that the ID-96 gene is conserved across all GBS serotypes (and strains) at both the gene and locus level.

APPENDIX I**ID-65****Forward Primer**5' - cggatccgccaccatgGCGGATCAAACTACATCGGTTC - 3'**Reverse Primer**5' - tgcggccgcGTTGGGATAACTAGTCGGTTAGTCG

Length (including restriction sites) = 1541bp

10 Incorporating 1515bp of gene-specific sequence encoding 505 amino acids of the putative mature protein.

Annealing temperature for PCR amplification = 60°C

Sequence predicted to encode a signal peptide was omitted from amplified product

15 **ID-66****Forward Primer**5' - cggatccgccaccatgAATCTTATTCCATAGTACTCCCTGC - 3'**Reverse Primer**5' - tgcggccgcAAAATGATCAGTTGAGGGTAAAGAG - 3'

20

Length (including restriction sites) = 767bp

Incorporating 747bp of gene-specific sequence encoding 247 amino acids of the putative mature protein.

Annealing temperature for PCR amplification = 60°C

25 Sequence predicted to encode a signal peptide was omitted from amplified product

APPENDIX II

ID-65**Forward Primer**5' - catgccatgCGGGATCAAACATACATCGGTT - 3'**5 Reverse Primer**5' - ttgcggccgcGTTGGGATAACTAGTCGGTTAGTCG

Length (including restriction sites) = 1534bp

Incorporating 1515bp of gene-specific sequence encoding 505 amino acids of the
10 putative mature protein.

Annealing temperature for PCR amplification = 60°C

ID-83**15 Forward Primer**5' - catgccatggcaAAAATAGTAGTACCAAGTAATGCCTC - 3'**ReversePrimer**5' - ttgcggccgcCTCTGAAATAGTAATTGTCCG - 3'

20 Length (including restriction sites) = 626bp

Incorporating 624bp of gene-specific sequence encoding 208 amino acids of the
putative mature protein.

Annealing temperature for PCR amplification = 52°C

25

ID-89**Forward Primer**5' - catgccatgggaAAGAAAGCAAATAATGTCAGTCC - 3'

Reverse Primer

5' - ttgcggccgcATTGGGTGTAAGCATTTC -3'

Length (including restriction sites) = 990bp

Incorporating 969bp of gene-specific sequence encoding 323 amino acids of the

5 putative mature protein.

Annealing temperature for PCR amplification = 54°C

ID-93**Forward Primer**

10 5' - catccatgggaACTGAGAACTGGTTACATACTAAAG - 3'

ReversePrimer

5' - ttgcggccgcATTAGCTTTCAACAATTCTC - 3'

Length (including restriction sites) = 759bp

Incorporating 744bp of gene-specific sequence encoding 248 amino acids of the

15 putative mature protein.

Annealing temperature for PCR amplification = 51°C

ID-96**Forward Primer**

20 5' - ctagctagccgATGTTGCGTGGAAAG - 3'

ReversePrimer

5' - ttgcggccgcATAAGATTAACAATACCAAGTAATATAGC - 3'

Length (including restriction sites) = 944bp

25 Incorporating 921bp of gene-specific sequence encoding 307 amino acids of the

putative mature protein.

Annealing temperature for PCR amplification = 53°C

rib (control)

Forward primer

5' - ggggtaccggccaccATGGCTGAAGTAATTCAAGGAAGT -3'

5 Reverse primer

5' - cggaattccgTTAACCTCTTTTTCTTAGAAACAGAT

Length (including restriction sites) = 3559bp

Incorporating 3531bp of gene-specific sequence encoding 1177 amino acids of the
10 mature protein.

Annealing temperature for PCR amplification = 55°C

APPENDIX III15 Listed below are the details (serotype and strain designation) of Group B
Streptococcus strains whose DNA was analysed for gene conservation

	SEROTYPE	STRAIN
20	Ia	515
	Ia	A909
	Ib	SB35
	Ib	H36B
	II	18RS21
25	II	1954/92
	II	118/158
	II	97/0057
	III	BM110
	III	BS30
30	III	M781
	III	97/0099
	IV	3139

V	1169/NT
VI	GBS VI
VII	7271
VIII	JM9

5

A group A Streptococcal strain (serotype M1, strain NCTC8198) and *Streptococcus pneumoniae* (serotype 14) were also included in the analysis for control purposes.

CLAIMS

1. A Group B Streptococcus polypeptide or protein having a sequence selected
5 from those described in fig 1, or fragments or derivatives thereof.
2. Derivatives or variants of the proteins, polypeptides, and peptides as claimed
in claim 1 which show at least 50% identity to those proteins, polypeptides and
peptides claimed in claim 1.
10
3. A Group B Streptococcus polypeptide or protein, or derivative or variant
thereof, as claimed in claim 1 or claim 2 , which is isolated or recombinant.
4. A nucleic molecule comprising or consisting of a sequence which is:
15
 - (i) any of the DNA sequences set out in figure 1 herein or their RNA
equivalents;
 - (ii) a sequence which is complementary to any of the sequences of (i);
 - (iii) a sequence which codes for the same protein or polypeptide, as those
sequences of (i) or (ii);
20
 - (iv) a sequence which shows substantial identity with any of those of (i), (ii)
and (iii); or
 - (v) a sequence which codes for a derivative, or fragment of a nucleic acid
molecule shown in figure 1.
- 25
5. A vector comprising one or nucleic acid molecules as defined in claim 4.

6. A vector as claimed in claim 4 further comprising nucleic acid encoding any one or more of the following: promoters, enhancers, signal sequences, leader sequences, translation start and stop signals, DNA stability controlling regions, or a fusion partner.

5

7. The use of a vector as claimed in claim 5 or claim 6 in the transformation or transfection of a prokaryotic or eukaryotic host.

8. A host cell transformed with a vector as defined in claim 5 or claim 6..

10

9. A process for producing a Group B Streptococcus polypeptide or protein, or derivative or variant thereof, as claimed in claim 1 or claim 2, the process comprising expressing the polypeptide or protein in a host cell as claimed in claim 8.

15

10. An antibody, an affibody, or a derivative thereof which binds to one or more of the proteins, polypeptides, peptides, fragments or derivatives thereof, as defined in any one of claims 1 to 3.

20

11. An immunogenic composition comprising one or more of the proteins, polypeptides, peptides, fragments or derivatives thereof as defined in any one of claims 1 to 3.

25

12. An immunogenic composition as claimed in claim 11 wherein the proteins, polypeptides, peptides, or fragments or derivatives thereof include ID-65 or ID-83, ID-89, ID-93 or ID-96.

13. An immunogenic composition as claimed in claim 11 or claim 12 which is a vaccine.

14. An immunogenic composition comprising one or more of the nucleic acid sequences as defined in claim 4.
- 5 15. An immunogenic composition as claimed in claim 14 wherein the nucleic acid sequences include ID-65 or ID-66.
16. An immunogenic composition as claimed in claim 14 or claim 15 which is a vaccine.
- 10 17. Use of an immunogenic composition as defined in any one of claims 11 to 16 in the preparation of a medicament for the treatment or prophylaxis of Group B Streptococcus infection.
- 15 18. A method of detection of Group B Streptococcus which comprises the step of bringing into contact a sample to be tested with at least one antibody, affibody, or a derivative thereof, as defined in claim 10.
- 20 19. A method of detection of Group B Streptococcus which comprises the step of bringing into contact a sample to be tested with at least one protein, polypeptide, peptide, fragments or derivatives as defined in any one of claims 1 to 3.
- 25 20. A method of detection of Group B Streptococcus which comprises the step of bringing into contact a sample to be tested with at least one nucleic acid molecule as defined in claim 4.
21. A kit for the detection of Group B Streptococcus comprising at least one antibody, affibody, or derivatives thereof as defined in claim 10.

22. A kit for the detection of Group B Streptococcus comprising at least one Group B Streptococcus protein, polypeptide, peptide, fragment or derivative thereof as defined in any one of claims 1 to 3.

5

23. A kit for the detection of Group B Streptococcus comprising at least one nucleic acid molecule as defined in claim 4.

24. A method of determining whether a protein, polypeptide, peptide, fragment or derivative thereof as defined in any one of claims 1 to 3 represents a potential anti-microbial target which comprises inactivating said protein and determining whether Group B Streptococcus is still viable.

10

FIG. 1

ID-65

Clone 3-60

GTGTTATGATGAAAAAAGGACAAGTAAATGATACTAAGCAA
TCTTACTCTACGTAAATATAAATTGGTTAGCATCAGTAA
TTTAGGGTCATTATAATGGTCACAAGTCTGTTTGCAGGA
TCAAACTACATCGGTCAAGTTAATAATCAGACAGGCACTAG
TGTGGATGCTAATAATTCTCCAATGAGACAAGTGCCTCAAGT
GTGATTACTTCCAATAATGATAGTGTCAAGCGTCTGATAAAG
TTGTAATAGCTAAAATACGGCAACAAAGGACATTACTACTC
CTTAGTAGAGACAAAGCCAATGGGAAAAAACATTACCTG
AACAAAGGAATTATGTTATAGCAAAGAAACCGAGGTGAAAA
ATACACCTTCAAAATCAGCCCCAGTAGCTTCTATGCAAAGAA
AGGTGATAAAGTTTCTATGACCAAGTATTAAATAAGATAAT
GTGAAATGGATTTCATATAAGTCTTGGTGGCGTACGTCGAT
ACGCAGCTATTGAGTCACTAGATCCATCAGGAGGTTCAGAGA
CTAAAGCACCTACTCCTGTAACAAATTCAAGGAAGCAATAATC
AAGAGAAAATAGCAACGCAAGGAAATTACATTTCACATA
AAGTAGAAGTAAAAAATGAAGCTAAGGTAGCGAGTCCAACTC
AATTACATTGGACAAAGGAGACAGAATTTCACGACCAAA
TACTAACTATTGAAGGAAATCAGTGGTTATCTTATAAATCATT
CAATGGTGTTCGTCGTTTGCTAGGTAAGCATCTTCA
GTAGAAAAAAACTGAAGATAAGAAAAAGTGTCTCCTCAACCA
CAAGCCCGTATTACTAAAACGGTAGACTGACTATTCTAACG
AAACAACACTACAGGTTTGATATTAAATTACGAATATTAAAGA
TGATAACGGTATCGCTGCTGTTAAGGTACCGGTTGGACTGAA
CAAGGAGGGCAAGATGATATTAAATGGTATACAGCTGTAAC
ACTGGGGATGGCAACTACAAAGTAGCTGTATCATTGCTGAC
CATAGAAATGAGAAGGGCTTATAATATTCACTTACTACC
AAGAAGCTAGGGACACTTGTAGGTGTAACAGGAACCTAAAG
TGACAGTAGCTGGAACTAATTCTCTCAAGAACCTATTGAAAA
TGGTTACCAAAAGACTGGTGTATATAATTATCGGAAGTACT
GAAGTAAAAAAATGAAGCTAAATATCAAGTCAGACCCAATT
ACTTTAGAAAAAGGTGACAAATAAATTATGATCAAGTATTG
ACAGCAGATGGTACCGAGTGGATTCTTACAAATCTTATAGTG
GTGTCGTCGCTATATTCTGTGAAAAAGCTAACTACAAGTAG
TGAAAAAGCGAAAGATGAGGGCACTAAACCGACTAGTTATCC
CAACTTACCTAAAACAGGTACCTATACATTTACTAAAACGTG
GATGTGAAAAGTCAACCTAAAGTATCAAGTCCAGTGGAAATT
AATTTCAAAAGGGTAAAAAATACATTATGATCAAGTGTAA
GTAGTAGATGGTCATCAGTGGATTCTACACAAGAGTTATTCCG
GTATTGTCGCTATATTGAAATTAA

MFMMKKGQVNDTKQSYSRKYKFGLASVILGSFIMVTSPVFADQTTSVQVNN
 QTGTSDANNSSNETSASSVITSNNDSVQASDKVVNSQNTATKDITPLVETK
 PMVEKTLPEQGNVYVSKETEVKNTPSKSAPVAFYAKKGDKVFYDQVFNKDN
 VKWISYKSFGGVRRYAAIESLDPGGSETKAPTPVTNSGSNNQEKIATQGNYT
 FSHKVEVKNEAKVASPTQFTLDKGDRIFYDQILTIEGNQWLSYKSFNGVRRFV
 LLGKASSVEKTEDKEKVSPQPQARITKTGRLTISNETTGFDILITNIKDDNGIA
 AVKVPVWTEQGGQDDIKWYTAVTTGDGNYKVAVSFADHKNEKGLYNIHLY
 YQEASGTLVGVGTGKVTVAGTNSSQEPIENGLPKTGVYNIIGSTEVKNEAKISS
 QTQFTLEKGDKINYDQVLADGYQWISYKSYGVRRYIPVKKLTSSEKAKDE
 ATKPTSYPNLPKTGTYTFTKTVDVKSQPKVSSPVÉNFQKGEKIHYDQVLVVD
 GHQWISYKSYSGIRRYIEI*

Sequence description

A) Length: 1642 bp - 547 aa (full length gene)

B) Sequence Characteristics:

Potential leader peptide sequence

Orf is preceded by a potential Shine-Dalgarno sequence.

ID-66

Clone 3-5

ATGATATTGAGACGTCGAACATTGTTTATGGCAACTGGGTATGCCATT
 TCTCTCATTCTTAGTATTCTAGCCTAAATCTTATTCCATAGTACTCCCTT
 GCAAACCAATGCAGCTTACGGAACCTTGCTCCCTCATTAACCACATCTTTT
 GGGACAGATGGTTAGGTAGGGATATGTTGTCAGAACGATTAAAGGACT
 TTATTCTCTACAAGTCGGCTTATTAGGTGCCCTATGGGGGTCTTCTG
 GCGACAGTTTGGAGTGCTGCAGGTTAGGAAATAGCATTATTGATAAA
 ATAATAGCATGGTTAGTTGATTGTTATTGGTATGCCATTGATTITA
 TGATTCTCATTCTTGTGTTGGAAAGGTGCTCAAGGGTCATCATTGC
 AACGGCTGTTACACATTGGCCTTCTTAGCAAGGCTTATCCGCAATGAAGT
 CTATCATCTAAAGAATAAAGAATTGTCACATTCTAAAGTATGGAAA
 AACGCCTTATTATATTGTGAGGCATCATATCCTGCCATTGATTGCTCTCAA
 ATTTCATTGGTTTATCCTCTTATTCCACATGTCTCATCCTACATGAAGCAT
 CAATGACTTTCTTAGGATTGGCTCTGCCAACAAACCTCGGTGGTA
 TCATTCTGTCAGAGGCAGCTAACATCTCTTGGAAATTGGTGGTGG
 TTATCTTCCAGGACTTATCTTATTGGTTGTCAATGCATTGATACTAT
 CGGAGAATCTTAAAGAAACTCTTACCCCTCAAACGTATCATTTAG

FIG. 1 _{CONT'D}

MILRRRTIVLWQLGIAISLILSILALNLYFHSTPLQTNAALRNLA
PSLNHLFGTD GLGRDMFVRTIKGLYFSLQVGLGALMGVILATVFGVL
AGLGNSIIDKIIAWL VDLFIGMPHLIFMILISFVVVGKGAQGVIIATA
VTHWPSLARLIRNEVYHLKNKE FVQLSKSMGKTPYYIVRHILPLIASQI
FIGFILLFPHVILHEASMTFLGFGLSAE QPSVGIILSEA
AKHISLGNWWLVIFPGLYLILVVNAFDTIGESLKLFYPQTDHF*

Sequence description

- A) Length: 822 bp - 274 aa (full length gene)
- B) Sequence Characteristics:
 - Potential leader peptide sequence
 - Orf is preceded by a potential Shine-Dalgarno sequence.

ID-78

Clone 3-5b

ATGACAGAAACATTATTAAGCATTAAAGACCTCTCCATCACCTCACTCAA
TACGGAAAGATTTAAAACCATTCAATCAACACCGATAACAAGCGCTGA
ATTAGAAATTAAAAAAGGTGAGTTATTAGCTATTATAGGTGCTAGTGGTT
CGGGGAAGAGTTATTAGCACATGCTATTATGGATATTCTCCTAAAAATG
CATCTGTAAACAGGAGATATGATTATCGTGGTCAATCACTAAATTCTAAC
GCATTAAACAGTTGCGAGGAAAAGATATTACGTTGATTCCACAATCAGTTA
ATTATTTAGATCCATCTATGAAAGTCACATCAGGTGCGCTTAGGTATCT
CAGAAAATTCAAAGGCTACTCAAGAAGGATTGTTCAACAGTTGGTTAA
AAGAAAAGTGATGGTGAATTGGATCCTTCCAACCTTCTGGCGGAATGCTCC
GACGTGTTTGTACACGTGATTAGTGATAAGGTTCTTGATTATTGC
GGATGAGCCCACCCCTGGATTACATCCAGATGCTCTGCAAATGGTTTAA
CCAAC TACGCTCCTTGCAGATAAAGGAATAAGCGTTATATTACTCA
TGATATTGTAGCAGCTAGTCAAATTGCTGATCGTATTACTATTAAAGA
GGGAAAAGCTATTGAAACAGCTCCAGCTAGTTCTTAGCGGAAATGGAG
AGCAGTTACAAACAGAAATTGCTAGAAGTTATGGCGCTCTCCCCACAGC
AAGAATTTTGAAAGGAGTTACTCATGACCTTAGAGGCTAA

MTETLLSIKDSLITFTQYGRFLKPFQSTPIQALNLEIKKGELLAIIGASGSGKSLL
AHAIMDILPKNASVTGDMIYRGQSLNSKRIKQLRGKDITLIPQSVDLDP
VKHQVRLGISENSATQEGLFQQFGLKESDGDDLPFQLSGGMLRRVLFTTCIS
DKVSLIIADEPTPGLHPDALQMVLQLRSFADKGVISVIFITHDIVAASQIADRITI
FKEGKAIETAPASFFSGNGEQLTEFARSLWRSLPQQEFLKGVTHDLRG*

FIG. 1 CONT'D

Sequence description

A) Length: 804 bp - 268 aa (full length gene)

B) Sequence Characteristics:

No obvious leader peptide sequence

Orf is preceded by a potential Shine-Dalgarno sequence.

This gene was not isolated using the LEEP system. However in determining a full length gene sequence for ID-76, this gene was identified downstream and fully sequenced.

ID-79

Clone 3-5c

GTCCATCTGGGGTGGTCCGATTGGTATTCTTCTCCGATAGGTACTTGAGTCAAGATATTACGTTAGCTGATCGTATTAAGCACCTTATTACCTGTTTACCGGTAAGTATTCTAGGCATTGCCAATGTAACCTCTCATACTAGAACTAAATGATGTCGGTACTTCTAGTGAATATGCTTATTGCCAGAGCGCGTGGGAAACGGAATGGCAAATTAAAATCATTGTCTTAGAAATGCTATCGTACCAAGCTATTACACTGCATTTCCTATTGGAGAATTGTTGGAGGATCCGTTCTGCTGAGCAAGTTCTCATATCCAGGACTAGGGTCTACCCTAACTGAAGCAGGACTAAAAGTGATACACCGCTACTCTAGCTATTGTGATGATAGGACATTATTGTTTGCGGGCAATCTATTGCGGATATTAAATAGCATAATCAATCCACAGTTAAGGAGAAAAGTATGA

VHLGWFPIGISSPIGTLSDITLADRIKHLILPVFTVSILGIANVTLHTRTKMMSVLSEYVLFARARGETEWQIFKNHCLRNAIVPAITLHF SYFGE LF GG SV LA EQ VF SYPGLGSTLTEAGLKSDTPLLAIVMIGTLFVFAGNLIADILNSIINPQLRRKV*

Sequence description

A) Length: 495 bp - 165 aa (partial gene sequence)

B) Sequence Characteristics:

N-terminus has yet to be determined.

This gene was not isolated using the LEEP system. However in determining a full length gene sequence for ID-76, this gene was identified upstream.

FIG. 1 CONT'D

ID-80

Clone 2-17

TTGCGGACAATTACGTTAAACACAATGAAACCGCAGTCGTAAAAAGCGA
 AGGTAGGGCGGTAATGCTAAAAGATTATTACTGAAGATGGGAATTGA
 CAAAGATTAGTCGTCGTTCTGGATGTTAGTGGTTATCTATTGTCTTAT
 TATTGTCAGGATGTGTTGGCCTCAAATTATGATTGAGGGGTATCAAC
 TCCGAATGTTCAGCGCTTCGGAAGAATTGAGCTCTTAGTACCATTTAA
 TTCTTTCTGTAGTTAGATCAGCTAACTAGCTTAAAGAGATTTTGGGTT
 ATTGGTCAAAATGTAGTGAATATTTACTGCTGTTCCCTCATTATAGGGT
 TACTATCCCTAAAGCCAAGTTACGGAAATATAAAAGCGTTATATTACTTG
 CTTCCTGATGTCTCTTTCATAGAGTGTACTCAAGTTGTTAGATATTT
 AATAGATGCTAACGGGTTTGAAATCGACGATCTATGGACAAATACCTT
 AGGCCTCCCTTCGCCCTATGGAGTTATCGAAACATAAAAGGTGGCTTCT
 AACTATTAGAAAATGA

MRTITFKHNETRSSKSEGRAVMLKRLFTEDGEELTKISRRFVWMLVVIYCLIVR
 MCFGQIMIEGVSTPNQRFRIVALLVPFNSFRSLDQLTSFKEIFWVIGQNVV
 NILLFPLIIGLLSLKPSLRKYKSVILLAFLMSLFIECTQVVLDILIDANRVFEIDD
 LWTNTLGGPFALWSYRNIKGWLLTIRK*

Sequence description

- A) Length: 579 bp - 193 aa (full length gene)
- B) Sequence Characteristics:
 - Possesses a potential leader peptide sequence
 - No obvious Shine-Dalgarno, but the 'TTG' codon may not be the actual translation start point.
 - A methionine (ATG) that occurs ~22 codons downstream of the 'TTG' is preceded by a potential Shine-Dalgarno sequence and may represent the actual start codon.

ID 81

Clone 3-1

FIG. 1 CONT'D

TTGAAAAATTAAATCGTTATGTAGTTGGCGTTCTGGAGTCGTTTACATT
TAATGCTAGGATCAACTTATGCTTGGAGTGTGTTCTGAACCCAATTATCT
CAGAGACTGGTGGGATATTCTCATCAGTTCTCGCTTTAGTTGGCTAT
TTTTGTCTAGGAATGTCTGCAGCTTTATGGGACACTAGTAGAGCGTTT
GGTCCTAGGATAATGGGAATGATTCTGCTATTATGGAGCAGGGAAAT
GTGTTAACAGGCTTAGCCATTGAAACTCAGCAGTTATGGTTACTGTATGTT
GCATACGGTATTTAGGAGGAATCGGACTTGGTCAGGTTATTAACCTCA
GTATCGACTATTATAATGGTTCTGATAGGAGGGACTAGCAACAGG
ATTGCTATTATGGGATTGGCTTGCTTCTTAGTAACAAAGTCCGCTTGCA
CAATCCCTACTGATTAGGATTGGTGTGGGAAAACGTTTATTTGGGA
TTAGTATATTGTCATGATGATTGCTCACAAATTATAAAACAACCAC
CTCAGGAAAAATAACTATTGACTCACGATGGTAAAAGAATGCTATG
AATTACAAATTATCACTGGATTAAAAGCAAACGTCGCTATAAAATCAA
AACCTTTACATCATTGGTTGACCTGTTATTAAATTTCGTGTGGCTTA
GGTTAATATCAGCAGCTCACCAATGGCACAAGATTAGCAGGCTATTCC
GCAGAACTGCAGCCTATTAGTAGGGGTACTAGGGATATTAAACGGTTT
GGACGTCTGTTATGGCAAGTCTCTGACTACATTGGACGCCGTTGACC
TTTATAATTATTTATGGAACCTTATTATGACTCTAGTTATTGTC
ATTCAATGCTATTGTATTGCAATAGCGATGCTATTAAATGACTTGTAT
GGTGCAGGTTTCTTATTACCTGCTTCTAAGTGTATTTGGAACAA
AGGAATTAGCTACTTACATGGTTATAGTTAACAGCATGGCAATAGCAG
GTCTGTTGGCCCCTATTGTTCAAAAGACATATTCAATGGGAAATTCT
ATCAATTGACATTAATGGTTTTGGTTTATTCTTATTGGATTATTGTTA
TCTCTATATTAAAGAAAATTAAACAACAAAGTTGTGTAG

LKNLNRYVVAVSGVVLHMLGSTYAWSVFRNPIISETGWDISSLFAFSLAIFC
LGMSAAFMGHLVERFGPRIMGMISAILYGAGNVLTGLAIETQLWLLYVAYG
ILGGIGLGSYITPVSTIIKWFPDRRGLATGFAIMGFGFASLVTPLAQSLIRIG
VGKTFYILGLVYFFVMMIASQFIKQPPQEKITILTHDGKKNAMNSQIITGLKAN
VAIKSKTFYIWLTLFNISCGLGLISAASPMAQDLAGYSAESAALLVGVLGIN
GFGRLLWASLDYIGRPLTFIILFIVNFIMTSSLFLSFNAIVFAIAMSILMTCYGA
GFSLLPAYLSDIFGTKELATLHGYSLTAWAIAGLFGPLLSKTYSWGNSYQLTL
MVFGFLFLFGLLLSSLYRKLTTKVV*

Sequence description:

- A] Length 1221 bp - 407 a.a (full length gene).
- B] TTG start codon with Shine-Dalgarno sequence upstream. Obvious signal peptide, with hydropathy plot exhibiting many possible membrane spanning regions, indicating protein to be transmembrane.

FIG. 1 CONT'D

ID-82

Clone 48

ATGGCAGATAAAACAGAACATTAAACTTGTAGGTGCAGGATCTCTAG
CACACAAGAAAAAATTGAAAAGCCTGCTCTTCGTTATGCAAGATGCGTG
GCGTCGCTGAAAAAAAACAAATTAGCAGTAGTTCACTCTATTATTAGC
TCTTTACTTACTTTCTGTTAGCCTCAAATTATTGTAACTCAGAAGGAT
GCTAATGGGTTTGAATCGAAAAAAAGTAACGACATATCGCAACTTACACCT
AAATTGAGTTCAAACCTCCCTTTGGAATGGTAGCATTAATCCATCA

MADKNRTFKLVGAGSSSTQEKIEKPALSFMQDAWRRLKKNKLAVVSLYLLA
LLLTFSLASNLFVTQKDANGFDSKKVTTYRNLPKLSNLFWNGSINPS

Sequence description:

A] Current length is 303 bp - 101 aa
B] No obvious signal peptide but Shine
Dalgarno sequence upstream of the ATG start
codon. Not identified directly using the LEEP system but was found
directly downstream of ID-34 described in WO 00/06736.

ID-83

Clone 98

ATGAAAATAGTAGTACCACTGAAATGCCCTCGCAGTCTGAAGAGGGCTCAAGA
AATAGATTATCAAAATTGATAGTGTGATATTATTGAATGGCGAGCTGA
TGCCTTACCAAAGGATGACATTATTAAATGTAGCTCCAGCTATTTGAGAA
ATTCGCAGGTATGAAATTATTACTTTCTGACAACGCGTGAAGGTGG
TAATATTGCTTATCTGATGCTGAGTATGTGAGTTAATCCAGAAAATTAA
TTCTATCTACAATCCAGATTATATTGATTTGAGTATTTTCACATAAAGAA
GTTTTCAAGAAATGCTAGAATTCCAATTAGTCCTGTCTTATCACAATT
TTCAAGAGACACCGGAGAATTATGGAGATATTTCAGAATTAAACAGCC
CTAGCACACGAGTTGTAAAATCGCAGTAATGCCAAAGAATGAACAAGA

FIG. 1 CONT'D

8 / 110

TGTCTTAGACGTTATGAATTACACTCGCGGTTCAAGACTATTAATCCTGA
 TCAAGTTTATGCGACGGTATCTATGAGAAAATTGGACGTATTCTCGTTT
 GCTGGTGATGTACTGGATCTAGTGGACATTGCATATTAGATTCATCT
 ATCGCACCCGGACAAATTACTATTTCAGAGATGAAGCGTGTCAAAGCATT
 GCTTGACGCTGACTGA

MKIVVPVMPRSLEEAQEIDLSKFDSVDIIEWRADALPKDDIINVAPAIFEKFAG
 HEIIFTFRRTREGGNIVLSDAEYVELIQKINSIYNPDYIDFEYFSHKEVFQEMLEF
 PNLVLSYHNFQETPENIMEIFSELTALAPRVVKIAVMPKNEQDVLDVMNYTRG
 FKTINPDQVYATVMSKIGRISRFAGDVTGSSWTFAYLDSIAPGQITISEMKRV
 KALLDAD*

Sequence description:

- A] Length 678 bp, 225 aa (full length gene)
- B] No obvious signal peptide, but there is a Shine Dalgarno immediately upstream of ORF.

ID-84

Clone RS-52

ATGAAAGACTTATTGCAACAAACAGAACAGCATCATCAAGGAAACAGGAACA
 AGATAGAATTGTCATTACATAAAACATGTTGAGTTAACAAATGGTA
 ATCAAATAAAAAAAATTGAGTTATCGACTTTCAAAAAATGAGATGACA
 GGTACATGGGAATTCTACTAAAATTGAACAATTTCGATTAGTTT
 TCTGAAGATAGAATTGGTGGTAAACTTAGAGCATTAGGATATCAACCGAA
 TGAAATAGGTTTCAAGGACATCAATAGTAATAATCAAAATGTTAATGA
 TATTGAAGTGATTATATGAAGAAAGAATAG

MKDLFATTEASSRKQEQRIVNYIKQHVELTNGNQIKKIEFIDFQKNEMTGTW
 GISTKINEQFSISFSEDRIGGKLRALGYQPNEIGFSKDINSNNQNVDIEVIYMK
 KE*

Sequence description:

- A] length: 333 bp - 111 aa (partial sequence)
- B] No obvious Shine Dalgarno sequence upstream
 of the ATG start codon, and no obvious signal
 peptide within the protein.

FIG. 1 CONT'D**SUBSTITUTE SHEET (RULE 26)**

ID-85

Clone RS-53

ATGAAAAAACGTATATGGTATTGATAATAATAATCACAGTAATTAGGA
GGACTAGCCATGAAAAACTATTGCAACAAACAGAAGCATCATCAAGGAA
ACAGGAACAAGATAGAATTGTCAATTACATAAAACAACATGTTGAGTTAA
CAAATGGTAATCAAATAAAAAAATTGAGTTATCGACTTCAAAAAAAT
GAGATGACAGGTACATGGGAATTCTACTAAAATTAAATGAACAATTTCG
ATTAGTTTCTGAAGATAGAATTGGTGGTAAACTTAGAGCATTAGGATAT
CAACCGAATGAAATAGGTTTCAAAGGACATCAATAGTAATAATCA

MKKRIWYLIIITVILGLAMKNLFATTEASSRKQEQRIVNYIKQHVELTNGN
QIKKIEFIDFQKNEMTGTWGISTKINEQFSISFSEDRIGGKLRALGYQPNEIGFSK
DINSNNQ

Sequence description:

- A] Length: 351 bp - 117 aa (Partial sequence)
- B] Obvious signal peptide and Shine Dalgarno sequence upstream of the ATG start codon.

ID-86

Clone ID-74

ATGCTAAATCAATATGATTATATCGTTATTGGTGGAGGTAGT
GCAGGCAGTGGTACCGCTAATAGGGCAGCCATGTATGGAGC
AAAAGTCCTGTTAATTGAAGGTGGACAAGTAGGTGGAACCTTG
TGTAACTTAGGTTGTGTACCTAAGAAAATCATGTGGTATGG
TGCACAAGTTCTGAGACACTCCATAAGTATAGTTAGGTAA
TGGTTTGAAAGCCAATAATCTTAGTTTGATTTTACTACTCTA
AAAGCTAACCGCGATGCTTACGTGCAGCGGTCTAGACAGTCG
TATGCCGCTAATTGAGCGTAATGGGTCGAAAAGATTGAT
GGATTTGCTCGTTTATTGATAACCATACTATTGAAGTGAATG
GTCAGCAATATAAGCTCCTCACATTACTATTGCAACAGGTG

FIG. 1 CONT'D

GACACCCTCTTACCTGATATTATTGGAAGTGAACCTGGTG
AGACTTCTGATGATTTTTGGATGGGAGACCTTACCAAATTC
TATATTGATTGTTGGGGCGGGCTATATCGCGGCAGAACTTGC
TGGAGTGGTTAATGAATTAGGCGTTGAAACCCATCTGCATT
TAGAAAAGACCATATTCTACCGGGATTGATGACATGGTAAC
AAGTGAGGTATGGCTGAAATGGAGAAATCAGGTATCTCTT
ACATGCTAACCATGTACCTAAATCTCTAAACGCGATGAAGG
TGGCAAGTTGATTTGAAGCTGAAAATGGAAAACGCTTGT
CGTTGATCGTGTAAATATGGGCTATCGGCCGTGGACCAAATGT
AGACATGGGACTGAAAATACCGATATTGTTAAATGATAA
AGATTATATCAAAACAGATGAATTGAGAATACCTCTGTAGA
TGGCGTGTATGCTATTGGAGATGTTAATGGAAAATGCCCT
GACACCGGTAGCAATTGCAGCAGGTGTCGCTTATCAGAAAG
ACTTTTAATCATAAAGATAACGAAAAATTAGATTACCATAA
TGTACCTTCAGTTATTTACTCACCTGTAATTGGGACGGTA
GGACTTCAGAACAGCAGCTATCGAGCAATTGGAAAAGAT
AATATCAAAGTCTATACATCAACTTTACCTCTATGTATACGG
CTGTTACCAAGTAATGCCAAGCAGTTAAGATGAAGCTCATAA
CCCTAGGAAAAGAGGAAAAAGTTATTGGGCTCATGGTGTG
GTTATGGTATTGATGAAATGATTCAAGGTTTCAGTTGCTAT
CAAAATGGGGCTACTAAAGCAGACTTGATGATACTGTTGC
TATTCAACCAACTGGATCTGAGGAATTGTTACAATGCGCTA
A

MSNQYDYIVIGGGSAGSGTANRAAMYGAKVLLIEGGQVGGTC
VNLCVPKKIMWYGAQVSETLHKYSSGYGFANNLSFDFTTLK
ANRDAYVQRSRQSYAANFERNGVEKIDGFARFIDNHTIEVNGQ
QYKAPHITIATGGHPLYPDIIGSELGETSDDFFGWETLPNSILVG
AGYIAAELAGVVNELGVETHLAFRKDHLRGFDDMVTVSEVMAE
MEKSGISLHANHPKSLKRDEGGKLIFEENGKTLVVDRVIWAI
GRGPNVDMGLENTDIVLNDKYIKTDEFENTSDVGVYAIGDVN
GKIALTPVIAAAGRRLSERLFNHKDNEKLDYHNVPSVIFTHPVIG
TVGLSEAAIAEQFGKDNIKVYTSTFTSMYTAVTNRQAVKMCLI
TLGKEEKVIGLHGVGYGYGIDEMIQGFSVAIKMGATKADFDDTVAI
HPTGSEEFVTMR*

ID-87

FIG. 1 CONT'D

SUBSTITUTE SHEET (RULE 26)

Clone RS-55

ATGACAAAAAAACATCTAAAACGCTTGCCTGGCACTTACTACAGTATCA
GTAGTGACATACAGCCAGGAGGTATATGGATTAGAAAGAGAGGAATCGGT
CAAACAAGAACAAACCCAGTCAGCTTCAGAAGATGATTGGITCGAAGAAG
ATAATGAGAGGAAACAAATGTTCTAAAGAGAATTCTACTGTTGATGAA
ACAGTTAGTGATTATTTCTGATGGAAATAGTAATAACTCTAGTTCTAAA
ACCGAGTCAGTGGTAAGTGACCTAAACAAGTCCCCAAAGCAAAACAGA
GGTTACACAAGAAGCAAGCAATTCTAGTAATGATGCTAGCAAAGTAGAAG
TACCAAAACAGGATACAGCTCAAAAAAGGAAACTCTAGAAACATCAACT
TGGGAGGCAAAAGATTCGTAACTAGAGGGGATACTTAGTAGGTTTCA
AAATCTGGAATTATAAGTTATCTCAAACATCACACTGGTTTACCAAGT
CATGCAGCAGATGGAACTCAATTGACACAAGTAGCTAGCTTGCTTTACT
CCAGATAAAAAGACGGCATTGCAGAATATAACAAGTAGGCTAGGAGAAA
ATGGGAAACCGAGTCGTTAGATATTGATCAGAAGGAAATTATTGATGAG
GGAGAAATATTAATGCTTACCAAGTTGACTAAGCTTACTATTCAAATGGT
TATAAGTCTATTGGTCAAGATGCTTTGTGGACAATAAGAATATTGCTGAG
GTTAACCTCCTGAGAGTCTGAGACTATTCACTATGCTTTGCTCACA
TGCTTAAACAAAGTAAAGTTACCAAGATAACCTAAAGGTATTGGAGAA
TTAGCTTTTTGATAATCAGATTGGGTAAGCTTACTTGCCACGTCACT
TGATAAAATTAGCAGAACCGCTTCAAATCTAATCGTATTCAAACAGTTG
AATTGGGAAGTAAGCTTAAGGTTAGGAGAAGCAAGTTCAAGAT
AATAATCTGAGGAATGTTATGCTCCGGATGGACTTGAAAAAATAGAATC
AGAAGCTTTACAGGAAATCCAGGAGATGAACATTACAACAATCAGGTTG
TATTGCGACAAGGACAGGCCAAATCCACATCAACTTGCGACTGAGAAT
ACTACGTCAATCCGGACAAATCATTGTGGCGTGCAACACCTGATATGGAT
TATACCAAATGGTAGAGGAAGATTACCTATCAA AAAAATAGTGTAC
GGTTTCAAATAAAGGCTACAAAAGGTAAAGACGTAAATAAAAACCTTAA
AATTCCAAAACAACACAATGGTATTACTATTACTGAAATTGGTATAACGC
TTTCGCAATGTTGATTTCAAAGTAAACCTTACGTAAATATGATTGGA
AGAAATAAAGCTCCCTCAACTATTGGAAAATAGGTGCTTTGCTTTCA
ATCTAATAACTGAAATCCTTGAAGCAAGTGAAGATTAGAAGAGATTA
AAGAGGGAGCCTTATGAATAATCGTATTGGAACTCTAGACTTGAAAGAC
AAACCTATCAAATAGGTATGCTGCTTCCATATTAAATCATATTATGCC
ATTGTTCTCCAGAATCTGTACAAGAAATAGGACGTTAGCTTACGCTTCA
AATGGTGCCTCACCTTATGGAAATAAGGTAAAACAATTGGT
GAAATGGCTTTTATCCAATAAACTGGAAAGTGTAAATCTCTGAGCAA
AAACAATTAAAGACAATTGAGGTCCAAGCTTTGGATAATGCCCTAGT
GAAGTAGTCTTACGCCAAATTACAGACTATTGCTGAAGAGGGCTTCAAA
AGGAATCATTGAAAGAAGTGAAGGGTCACTACATTATCTCAGATTACT
TTAATGCTTTGATCAAATGATGGGGACAAACGCTTGGTAAGAAAGTG
GTTGTTAGGACACATAATAATTCTCATATGTTAGCAGATGGTAGCGTTT
ATCATTGATCCAGATAAGCTATCTTACAATGGTAGACCTTGAAAAGGTT

FIG. 1 CONT'D

TTAAAAATAATCGAAGGTTAGATTACTCTACATTACGTCAAGACTACTCAA
 ACTCAGTTAGAGAAATGACTACTGCAGGTAAAGCGTTATCAAATCT
 AACCTCCGACAAGGAGAAAAACAAAATTCTTCAAGAACACAATT
 CCTTGGTCGCGTTGATTGGATAAAGCCATAGCTAAAGCTGAGAAGGCTT
 AGTGACCAAGAAGGCAACAAAGAATGGTCAATTGCTGAGAGGAGTATT
 ACAAAAGCGGTATTAGCTTATAATAATAGTGCTATTAAAAAGCTAATGTTA
 AGCGCTTGGAAAAAGAGTTAGACTTGCTGACAGATTAGTCGAGGGAAAA
 GGACCATTAGCGCAAGCTACAATGGTACAAGGAGTTATTATTAAAGAC
 GCCTTACCATGCCAGAATATTATCGGATTGAACGTTATTGACAA
 GTCTGGAAAATTGATTATGCACTTGATATGAGTGTACTATTGGCGAGGG
 ACAAAAAGATGCATATGGTAATCCTATATTAAATGTTGACGAGGATAATG
 AAGGTTATCATACCTTGGCAGTTGCCACTTAGCTGATTATGAAGGTCTT
 ATATTAAAGATATTAAATAGTCCCTGATAAGATTAAAGCAATACGCC
 AGATTCCITTGGCAAAATATCATAGATTAGGAATTITCCAAGCTATCCGAA
 ATGCAGCGGCAGAAGCAGACCGATTGCTTCTAAGACACCTAAGGGTAC
 CTAATGAAGTCCAAATTATCGAAAAAAACAAATGGAGAAAATTAAA
 ACCAGTTGATTATAAAACGCCGATTTTATAAAGGCTTACCTAATGAAAA
 GGTAGACGGTGTAGAGCGGCTAAAGGTATAATATAATGCGGAGACTA
 ATAATTCTGTAGCTGTAACACCAATAAGGTCGAGCAGCAATTACATAAGT
 CACAGTCTGATGTAATTACCTCAAACAAGTTCTAAAATAATTATAT
 ACGAGATTCTAGGATACGTTAGTTATGTTGCTTCTAGTAACTGCTGG
 GAAAAAAGGAAAACGAGCAAGAAAATAA

MTKKHLKTLALALTTSVVTYSQEVLEREEVKQEQTQSASEDDWFEEDN
 ERKTNVSKENSTVDETVSDLFSDGNSSNSKTESVVSDFPKQVPKAKPEVTQE
 ASNSSNDASKVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKL
 SQTSHLVLP SHAADGTQLTQVASFAFTPDKTKTAIAEYTSRLGENGP
 KPSRLDIDQ KEIIDEGEIFNA YQLTKLTIPNGYKSIGQDAFVDNK
 NIAEVNLPELETISDYAF AHMSLKQVKLPDNLK VIGELAFFDNQIGGKLYLPRH
 LIKLAERA FKSNRQTV EFLGSKLKVIGEASFQDN
 NLRNVMLPDGLEKIESEAFTGNPGDEHYN
 NQVVLRTTGQNP
 HQLATENTYVNPD
 KSLWRATP
 DMDYTKWLEEDFTYQ
 KNSVTGFS NKGLQKVR
 RNKNLEIPKQHNG
 ITITEIGDNAFR
 NVDFQS
 KTLRKYD
 LEEIKLPS TIRKIGAF
 QSNNLKS
 FEASEDLEE
 IKEGA
 FMNNR
 IGTLDL
 KDKL
 KIGDAAFH
 INHIYAI
 VPESV
 QEIGRSA
 FRQNG
 ALHLM
 FIGNK
 VKTIG
 MAFLS
 NKLES
 VNL SEQKQL
 KIEV
 QAFSD
 NALSE
 EVV
 LPP
 NLQT
 IRE
 EAF
 KRN
 HLKEV
 KG
 S
 TLS
 QITF
 NA
 FD
 Q
 ND
 G
 DK
 RFG
 KK
 VV
 VR
 THNN
 SHML
 AD
 GER
 FI
 ID
 P
 D
 K
 LS
 S
 T
 M
 V
 D
 L
 E
 K
 VL
 D
 L
 L
 V
 E
 G
 K
 G
 P
 L
 A
 Q
 A
 T
 M
 V
 Q
 G
 V
 Y
 L
 L
 K
 T
 P
 L
 P
 E
 Y
 I
 G
 L
 N
 V
 Y
 F
 D
 K
 S
 G
 K
 L
 I
 Y
 A
 LD
 M
 S
 D
 T
 I
 G
 E
 G
 Q
 K
 D
 A
 Y
 G
 N
 P
 I
 L
 N
 V
 D
 E
 D
 N
 E
 G
 Y
 H
 T
 L
 A
 V
 A
 T
 L
 A
 D
 Y
 E
 G
 L
 Y
 I
 K
 D
 I
 L
 N
 S
 S
 L
 D
 K
 I
 A
 R
 Q
 I
 P
 L
 A
 K
 Y
 H
 R
 L
 G
 I
 F
 Q
 A
 I
 R
 N
 A
 A
 A
 E
 A
 D
 R
 L
 L
 P
 K
 T
 P
 K
 G
 Y
 L
 N
 E
 V
 P
 N
 Y
 R
 K
 Q
 M
 E
 K
 N
 L
 K
 P
 V
 D
 Y
 K
 T
 P
 I
 F
 N
 K
 A
 L
 P
 N
 E
 K
 V
 D
 G
 D
 R
 A
 K
 G
 H
 N
 I
 N
 A
 E
 N
 N
 S
 V
 A
 V
 T
 P
 I
 R
 S
 Q
 Q
 L
 H
 K
 S
 Q
 S
 D
 V
 N
 L
 P
 T
 S
 K
 N
 N
 F
 I
 Y
 E
 I
 L
 G
 Y
 V
 S
 L
 C
 L
 L
 F
 L
 V
 A
 G
 K
 G
 K
 R
 A
 R
 K
 *

FIG. 1 CONT'D

Sequence description:

- A] Length 3168 bp - 1056 aa (Partial sequence)
- B] Obvious signal peptide with Shine Dalgarno sequence upstream of the ATG start codon.

ID-88**Clone RS-56**

GCAGGATACATCATGCACAAGCACGAGGCTATCGTGTATGCTGGGTCA
ACCCAGGAAGACATGTCGGCACAAGCTGAAGATTCTTACAGTCTGTACA
CAATAAAGAGACGGGTAAGAGCGCTTTAATGACAAAGAACGACTAGCAA
TT

AGYIMHKHEAIIVSCWGQPRKTCRHKLKISLQSVHNKETGKSAFNDKERLAI

Sequence description:

- A] Length:153 bp - 51 aa (partial sequence)
- B] No signal peptide visible, insufficient sequence data to determine the presence of a Shine Dalgarno sequence.

ID-89**Clone RS-58**

GTGTCATTTATGCAAAGAAAATCCTATTTAAAATCCATGAGTGTTCTTA
TTAACAGCTTGTCTTATATCAGGATATGTGGTAAAGATATTGCTATGTTA
CATGCAGTATCTGCCAGTGAGAAGAAAGCAAATAATGTCAGTCCGAGAGA
AAATCTCTACAGGGCTGTCAATGATAATTGGCTAGCCAATACAAAATCA
AACAAAGGGCAGACTAGTGTAAATAGTTTCAGAAATTGAGGATAAATT
AAGCAACTGTTAGTGTCTGATATGGCTAAATGGCCTCAGGAAAGATTGA

FIG. 1 CONT'D

AACAACCAATGATGAACAGAAAAAAATGGTTGCATACTATAAACAGGTA
TGGACTTTAAAACAAGAGATAAAAATGGTCTCAAACCTCTAAAACCAGTT
TTACAAAAAACTGAAAGCAGTCTTCATGAAAGACTTCAAAGTTGCC
CATGATTGTGATGAGTGGTTGTTTACCATTTGGTTGACTGTGGAAA
CCAATGCTCGAGATAATAGCCAAAAGCAATTGGTGCCTCGTCAAGCACCC
GCATTACTTGAATCACCTGACCAATATAAGAAGGGCAATAAGAAGGTGA
GGCTAAATTATCAGCTTACCGTACTTCAGCAATGGCTTGCTTAAACAAGC
TGGAAAAAGTAACATTGAAGATAGAAAAGTAGTTAAACAAGCTATAGCAT
TTGATAGACTCTTATCAGAAAAAACGCAAGTTGATCAAAGTAAATCACA
GCTGAAAGTGAGACAGCTGGGGCGATATAACCTGAAAGTATGGAAAC
GGTCACAATTACGCCAAGGAATTGACTTAAAGAATTGATTGAAAAACT
AGTTGGGCCAACGAATAAGGCAGTCAATGTAGAAGATAAAACTTATTTA
AACAGGTTAATGATGTTATAAATAGTAAACAATTAGCCAATATGAAAGCA
TGGATGATGATTCTATGCTAGTTGATCAATCAGATTCTAGGAGAACAA
AATCGTCAAGCAGCGAGTGCTTTAAGAATGTTGCGTCTGGTTGACTCAG
ATTGAATCGAAAGAAAAATGCTTACACCCAATTAG

MSFMQRKSYLKSMsvLTLTACLISGYVVKDIAMLHAVSASEKKANNVSPREN
LYRAVNDNWLANTKLQGQTSVNSFSEIEDKLKQLLVSMDMAKMASGKIETTN
DEQKKMVAYYKQGMDFKTRDKNGLKPLPVLKLEAVSSMKDFQSLAHDF
VMSGFVLPPGLTVETNARDNSQKQLVLRQAPALLESQDQYKKGNKEGEAKLS
AYRTSAMALLKQAGKSNIEDRKLVKQIAFDRLSEKTQVDQSKITAESETA
GRYNPESMETVHNYAKEFDFKELIEKLVGPTNKAVNVEDKTYFKQVNDVINS
KQLANMKAWMMISMLVDQSDFLGEQNRQAASAFKNVASGLTQIESKEMLT
PN*

Sequence description:

- A] Length: 1095 bp - 365 aa (full length gene)
- B] an GTG (possible ATG start codon located 7 bp further downstream) start codon with an obvious signal peptide. Shine Dalgarno sequence present upstream of the ORF.

ID-90

Clone RS-59

FIG. 1_{CONT'D}

15 / 110

ATGGAAATGCCTAAAAGAAATGAATTACTCAATAAAGAAATTAAAATGAG
TATTGATAAACTTAGATATAAAGAACCGAGAGTGAAACATGACAAGCGAC
CTACTTTTATTGGTAGTACTTATACTGTTACTGTAGCAGTTATTTGTC
GTATTAAATATTTTATAG

MEMPKRNELLNKEIKMSIDKRYKEPESEHDKRPTFYLVVLILVTAVILSLFK
YFL*

Sequence description:

- A] Length: 174 bp - 58 aa(full length gene)
- B] No obvious signal peptide, but Shine Dalgarno sequence is present upstream of ATG start codon.

ID-91

Clone RS-62 (partial sequence)

ATGCAGGTATTTAAATATTGTCATAAAATTCTTGATCCAGTTATTCTATA
TGGGTTGGGAGTTGTGATGCTAATTGTCATGACAGGTTAGCCATGATAT
TTGGAGTGAAGTTCTAAAGCACTGAAGGTGGTAT

MQVFLNIVNKFFDPVIHMGSVVMLIVMTGLAMIFGVKFSKALEGG

Sequence description:

- A] Length: 141 bp - 41 aa (partial sequence)
- B] Shine Dalgarno sequence present upstream of ATG start codon with a possible signal peptide present

ID-92

FIG. 1 CONT'D

Clone RS-69 (partial sequence)

ATGAAAAAGAAAACATTCACTGCTTATAACTTTAACGGCTTATCCTT
TGTCTTTGACAGTGCTTTATCTTCCATTATTGGATTATGACAGGAG
CTTTAA

MKKKTFSAYNFLTALILCLLTVLFIFPFYWIMTGAF

Sequence description:

- A] Length: 110 bp -36 aa (Partial sequence)
- B] Possible signal peptide with Shine Dalgarno sequence directly upstream of the ATG start codon.

ID-93

Clone RS-70

ATGACTGAGAACTGGTTACATACTAAAGATGGTCAGATATTATTATCGT
GTCGTTGGTCAAGGTCAACCGATTGTTTTTACATGGCAATAGCTTAAGT
AGTCGCTATTTGATAAGCAAATAGCATATTTCTAAGTATTACCAAGTT
ATTGTTATGGATAGTAGAGGGCATGGCAAAAGTCATGCAAAGCTAAATAC
CATTAGTTCAAGGCAAATAGCAGTTGACTAAAGGATATCTTAGTTCAATT
AGAGATTGATAAAGTTATATTGGTAGGCCATAGCGATGGTCTAATTAGC
TTAGTTTCAAACGATGTTCCAGATATGGTAGAGGGCTTGTCTTAAT
TCAGGGAACCTGACTATTATGGTCAGCGATGGTGGGATATTCTTAGTA
AGGATTGCCTATAAATTCTTCACTATTAGGGAAACTCTTCCGTATATG
AGGCAAAAGCTCAAGTTATTCGCTTATGTTGGAGGATTGAAGATTAGT
CCAGCTGATTACAGCATGTCACCTCTGTAATGGTTTGGTGGAAAT
AAGGACATAATTAAAGTTAAATCATTCTAAGAAACTTGCTTCTTATTCCA
AGGGGGAGTTTATTCTTAGTTGGCTTGGCATCACATTATAAGCAA
GATTCCCAGTTTTAATATTATTGCAAAAAAGTTATCAACGATACTTG
AAAGGAGAAATTGTTGAAAAAGCTAATTGA

MTENWLHTKDGSIDIYRVVGQGPIVFLHGNLSRIFYFDKQIAYSKYYQVIV
MDSRGHGKSHAKLNTISFRQIAVDLKDLVHLEIDKVILVGHSDGANLALVFQ

FIG. 1 _{CONT'D}

TMFPDMVRGLLNSGNLTIHGQRWWDILLVRIAYKFLHYLGKLFYMRQKA
QVISLMLEDLKISPADLQHVSTPVMVLVGNKDIKLNHSKKLASYFPRGEFYSL
VFGFHHIIKQDSHVFNIIAKKFINDTLKGEIVEKAN*

Sequence description:

- A] Length: 744 bp - 248 aa (full length gene)
- B] No obvious signal peptide, but Shine Dalgarno sequence upstream of the ATG start codon.

ID-94

Clone RS-71

ATGGTAGAAAAGAGTTAGGTAAAAATAGCTTACTATCCAACTATTGT
TCTAATTGCTCCGCAGGTACTGCCATTGCAGTTGTATATAATGATGACCAT
TCTTCTTAAGATACGGCTATCCCGAGTCTCCACTTCATATTTTATCAATA
CACGGATCATTGCACAGGCACCAAGCAAATATTTTGGGCTGGTATTGGGG
ACGGTATTCAAAAGCCCCTGAAGTAGAACGTGCTACCTTAGAGGGCTAAG
ACCAATAAACTACCACATACTGCAGTGTAGGACAAGCAGTCGCTCTGTCT
TCAAAGGAAGCTTTATCAATTGGTGAACAAGGTCTAAAGACGTTGAA
GCTAATTAGCTCGCGTGCAGTTGAAGAAATTGCGCTTGATATCTTA

MVAKELGKNSFTIPTICSNCSAGTAIAVVYNNDDHSFLRYGYPESPLHIFINTRIIA
QAPSKYFWAGIGDGISKAPEVERATLEAKTNKLPTAVLGQAVALSSKEAFY
QFGEQGLKDVEANLASRAVEEIALDIL

Sequence description:

- A] Length: 405 bp - 135 aa (Partial sequence)
- B] No obvious Shine Dalgarno sequence upstream of the ATG start codon, probable signal peptide present at the N-terminus.

ID-95

FIG. 1 CONT'D

Clone RS-73

TTGAGGGAAACTTACTGGAAAATTCAAGCGATTGCATAAAATAATCTT
 GCAGAGTTCTAGAGAAAGGAGGTACAGATTATTGGAGTGGCAAGATCT
 AGCGCAGTTACCTGTATCTATTTAAAGACTATGTTACAGATGCTCAAGA
 CGCGGAAAAACCTTTATATGGACAGAAGTATTAAAGGGAGATTAATCG
 CTCAAATCAAGAAATTATTTCATATTGGCCGATGACTAAGACAGTCAT
 TCTGGGGATGTTAGATCGAGAATTACACATTAGAATTAGCTAAAAAG
 AAATCATCAGTCGTGGTTATGAACCAGTTGTTCGGAATTGGAGGTCTCG
 CAGTTGTAGCTGATGAAGGAATTAAATTTCATTGGTTATTCCAGATGT
 TTTGAGAGAAAATTGTCATCTCAGATGGGTATCTTATAATGGTCGATTT
 ATTAGAACGTTATTCGGATTTATCACACCTATTGAGCACTTGAGTA
 GAGACCTCTATTGTCCTGGTAAGTTGATCTTAGTATAATGGCAAAAAA
 TTTGCTGGCTTGGCTAGCGCCGTAAAGAATGGTATTGCGGTATCAATT
 TACCTTAGCGTTGGCGATCAAAAGGGCGGAGTCACATGATTTCAGAT
 TTTATAAGATTGGCTAGGTGATACGGGTAGTCCAATTGCTTATCCAAT
 GTAGATCCTGAAATTATGGCTAATCTATCTGATCTATTAGATTGCTTATG
 ACAGTAGAAGATGTTATTGATCGTATGTTGATTAGCCTAAACAAGTAGGT
 TTTAATGATCGTTACTGATGATTAGACCCATTAGTTGCAGAGTTGAT
 AGATTTCAGGCTAAGTCTATGGCTAATAAGGGATGGTGANAGAGATGA
 ATAA

MRETYWKISSDCDKINLAEFSRERRSDLLEWQDLAQLPVSIFKDYVTDAQDAE
 KPFIWTEVFLREINRSNQEILHIWPMTKTVILGMLDRELPHLELAKKEIISRGYE
 PVVRNFGGLAVVADEGILNFLVIPDVFERKLSISDGYLIMVDFIRSIIFSDFYQPI
 EHFEVETSYCPGKFDLSINGKKFAGLAQRRIKNGIAVSIYLSVCGDQKGRSQMI
 SDFYKIGLGDGTGSPAYPNVDPEIMANLSDLLDCPMTVEDVIDRMLISLKQVGF
 NDRLLMIRPDVAEFDRFQAKSMANKGMVRDE*

Sequence description:

- A] Length: 921 bp -307 aa (Full-length gene sequence)
- B] No obvious Shine Dalgarno sequence upstream
 of the TTG start codon or signal peptide
 visible. Actual start point may be a further
 85 bp downstream (TTG). This start point is
 preceded by a typical Shine-Dalgarno sequence.

FIG. 1 CONT'D

ID-96

Clone RS-74

TTGGAAGGTTACTTATTGCATTGATTCCCATGTTGCGTGGGAAAGTATT
 GGATTGTTAGTAATAAAATTGGAGGGCGTCAAATCAACAAACATTGG
 AATGACTTCTAGGAGCATTGCTATTGCGATTATCGTATGGTTATTAAACA
 GCCAGAGATGACTGCCTCATTGGATTITGGTATCTTAGGTGGTATCCT
 ATGGTCAGTCGGCCAAAATGGTCAATTCAAGCAATGAAATATATGGGAG
 TCTCTGTGCTAATCCACTGTCAAGTGGTGCACAATTAGTAGGTGGAAGCC
 TAGTTGGTGCCTTAGTCTTCATGAATGGACTAACCCAATTATT
 AGGATTGACAGCGTTGACATTATTAGTTATCGGCTCTATTCTCAAGTAA
 ACGTGATGTTTCAAAGACAGCTTGGCAACACATCAAGAGTTTCAAAAG
 GATTGCTACAATTGCTTATTCAACTGTAGGTTACATCTGTACGAGTTT
 ATTAAACAACATTATGAAGTTCGACGCTATGGCCGTCAATTACCCATGGC
 TGTTGGAATGTGCTAGGTGCAATTGTTCATGAAGTTCTGTGTTAATT
 GAGGCTGTTGTTAAAAATATGATTACAGGTCTCATGTGGGGCGTTGGT
 AATGTCTTCATGTTATTGGCAGCAGCTAACAGCAGGGCTAGCAATTGCTTT
 AGTTTTCTCAACTGGAGTAATTATCTCTATTATTGGTGGTATTATT
 AGGTGAGACAAAAACGAAGAAAGAGCAGAAATGGGTTGTATGGGTATC
 CTTGTTTGTATGGGTCTATTACTGGTATTGTTAAATCTTATTAA

MEGLLIALIPMFAWESIGFVSNKIGGRPNQQTFGMTLGALLFAIIVWLFKQPEM
 TASLWIFGILGGILWSVGQNGQFQAMKYMGVSVANPLSSGAQLVGGSLVGAL
 VFHEWTKPIQFILGLTALTLVIGFYFSSKRDVSEQALATHQEFSKGFAТИYST
 VGYISYAVLFNNIMKFDAMAVILPMAVGMCMLGAICFMKFRVNFEAVVVKNMI
 TGLMWGVGNVFMLLAAGLAGIAAFSFSQLVIIIGGILFLGETTKKEQK
 WVVVMGILCFVMGAILLGIVKSY*

Sequence description:

- A] Length: 867 bp - 289 aa (full-length gene)
- B] Possible Shine Dalgarno sequence upstream of GTG start codon, no obvious signal peptide present.

ID-97

FIG. 1_{CONT'D}

SUBSTITUTE SHEET (RULE 26)

Clone RS-75

ATGACAACTTACTACGAAGCTATAAAGTGGAAACGAAATTGAAGATGTTAT
 TGATAAAATCAACTGGAAAAACTAACCGAACAAATTGGCTCGATACAC
 GTATCCCTTATCAAATGACTAGACGATTGGCGAAACTTCCGCTCAAG
 AAAAAGATCTGGCAAGGTTTGAGGCTTAACCTACTTGATACCA
 TGCAATCAGAAACTGGTGTGAAGCTATTGTCGCCATGTCGCACGCC
 ACGAAGAAGCTGTCTAAACAATTCAATTGAAATCTGTTACGCTA
 AATCTTATTCTCAATTCTCAACTTAAATACTAAATCAGAAATTGAAG
 AAATTCGAGTGGACTAATAATAATGAGTTCTCAAGAAAAAGCACGT
 ATTATCAATGACATTATGCTAATGAAATGCCCTCAAAAAAGGTGGCT
 TCCACCTACCTCGAAACTTCTTTTATTCTGGCTTTACACCTCTTA
 CTATTGGGAAATAATAAGTTAGCAAATGTTGCTGAAATCATTAAATTAAAT
 TATTCTGTGATGAATCTGTACATGGTACTTATATCGGTTACAAATTCCAGCTT
 GGTTTAACGAATTACAGAAGATGAGCAAGAGAAATTCTGTGATTGGAT
 GTATGACCTCTTATCAGCTGTGAAACGAAGAAAAATACACCAAGA
 CACTTATGATGGCGTAGGATGGACTGAAGAAGTTATGACCTTTACGCT
 ACAATGCTAATAAAGCTTATGAATTAGGACAAGATCCTTATTCCAG
 ATACAGCAAATGATGTCAACCCAATTGTTATGAATGGTATTCAACAGGAA
 CATCAAACCATGACTTCTCTCAAGTAGGTAATGGTACCTACTGGTA
 GCGTTGAAGCTATGCATGATGACTATAACTATGGATTATAA

MTTYYEAINWNEIEDVIDKSTWEKLTEQFWLDTRIPLSNDLDDWRKLSAQEK
 DLVGKVFGLTLLDTMQSETGVEAIRADVRTPHEEAVLNNIQFMESVHAKSY
 SSIFSTLNKSEIEEIFEWTNNEFLQEKKARIINDIYANGNALQKKVASTYLETF
 LFYSGFFTPLYYLGNKLANVAEIIKLIIRDESVHGTYIGYKFQLGFNELPEDEQ
 ENFRDWMDLLYQLYENEKYTKLYDGVGWTEEVMTFLRYNANKALMNL
 GQDPLFPDTANDVNPIVMNGISTGTSNHDFFSQVGNGYLLGSVEAMHDDDYN
 YGL*

Sequence description:

- A] Length: 960 bp - 320 aa (full length gene)
- B] Shine Dalgarno sequence present upstream of ATG start codon, but no signal peptide present.

ID-98

FIG. 1 CONT'D

SUBSTITUTE SHEET (RULE 26)

Clone RS-77 (partial sequence)

ATGAATTGGTCACGTATCTGGAACTCGTAAAAATTAAATATCCTTATTCA
 AACCCCTCAGACTCTATCGGCACTAAGAAAAAGCAAGAAAAGCATCCTAA
 AAAAGAATTTCAGCTTATAAATCCATGTTAGAAATCAGTTATTCAGAT
 TTTGCTCTTCAATAATTATGTATTCTCTTGATCACITGATTAAAG
 AATATCCGGGCTATTACGTTCTACATTGGTATCTTACACTAGTATCCAT
 TATCTACTCTTTATTGCGATGTACAGTGTCTATGAGAGTGACGATGTT
 AA

MNWSRIWELVKINILYSNPQLSALRKKQEKPKEFSAYKSMFRNQLFQILL
 FSIIYVFLVSLDFKEYPGYFTFYIGIFLVSIIYSFIAMYSVFYESDDV

Sequence description:

- A] Length: 311 bp - 103 aa (Partial sequence)
- B] Shine Dalgarno sequence present upstream of ATG start codon, no obvious signal peptide at N-terminus.

ID-99

Clone RS-78 (partial sequence)

TAATCTTTAGTCAACGGAGCAACAGGAAAATTGCAGGCTATGCGACAGA
 TATTCCACCACATAATTAGCAGAAGTCATTGATGCTGCTGTACATGAT
 TGATCACCTAAAGCTAAATTAGATAAAATTAAATGGAATTCTACCTGGTCC
 AGATTITCCAAGTGGCGCTATCATTCAAGGAAAAGATGAAATTGTAAGG
 CATATGAGACTGGTAAGGGGAGAGTAGCGGTTCGCTCGCGAACTGCTATT
 GAAACCTAAAAGGTGGTAAGAAACAAATTATTGTTACTGAAATTCTTAT
 GAAAGTTAAT

SFSQRSNRKIAGYATDIPPHNLAEVIDAVVYMDHPKAKLDKLMEFLPGPDFPT
 GAIIQGKDEIRKAYETGKGRVAVRSRTAIETLKGGKKQIIVTEIPYEVN

Sequence description:

- A] Length: 312 bp - 104 aa (Partial sequence)
- B] No obvious Shine Dalgarno sequence or a

FIG. 1 CONT'D

signal peptide. Both N- and C- termini of ORF
yet to be elucidated.

ID-100

Clone RS-79

ATGGGACGTAAAGTGGCCAATATTGTTGCCAAAAAGACTGCTAAAGATGG
TGCTAACTCAAAAGTATACTGCTAAATTGGTGTGAAATATATGTTGCTGC
AAAGCAAGGTGAACCAGACCCCGAGTCAAACTCAGCTCTAAAATTGTT
TGGACCGTGCTAAGCAAGCACAAAGCATGTTATTGATAAAAGCG
ATTGATAAAAGCCAAGGAAACACAGATGAAACTTCTGTAGAGGGACGCTA
TGAAGGTTTGGTCAAATGGTCAATGATTATTGTGGATACTTGTACATC
AAATGTTAACCGTACGGCAGCAAATGTACGTACTGCTTACGGTAAGAACG
GTGGCAATATGGGAGCTTCAGGATCGGTATCCTACTTATTGATAAAAAAG
GTGTATCGTTTGCTGGTGTAGATGCTGACACTGTCTTCGAACAATTACT
TGAAGCGGATGTAGACGTAGATGTTGAAGCAGAAGAGGGAACAATA
ACAGTTATACCGCCCCAACAGATCTCATAAAAGTATCCAAGCACTCGC
GATAATGGTGTAGAAGAATTCCAAGTTACTGAACCTGAAATGATTCTCAA
TCAGAAGTAGTATTGGAAGGTGATGACCTGAAACTTTGAAAAGCTT

MGRKWANIVAKKTAKDGANSKVYAKFGVEIYVAAKQGEPDPESNALKFVL
DRAKQAQVPKVIDKAIDKAKGNTDETFVEGRYEGFGPNMSMIIVDTLTSNV
NRTAANVRTAYGKNGGNMGASGSVSYLFDKKGIVFAGDDADTVFEQLLEA
DVDVDDVEAEEGTITVYTAPTDLHKGIQALRDNGVEEFQVTELEMIPQSEVVL
EGDDLETFEKL

Sequence description:

- A] Length: 654 bp - 218 aa (Partial sequence)
- B] Possible Shine Dalgarno sequence upstream
of ATG start, no obvious signal peptide

ID-101

Clone RS-80

FIG. 1 CONT'D

23 / 110

TTGGAGAAATATTGAAGAACCGATTACATGGATTGGATTAGTTCTTGTG
GTTACGTGGTTTTAACTAAAAGTAGTGAATTGGATTTGGTGTGTG
TCTTGTGTTAGTATTGCTAGTCAGTCAAAAGTGAT

MEKYLKNPITWIGLVVVTVFLTKSSEFLIFGVCVLLLVFASQSD

Sequence description:

- A] Length: 135 bp - 45 aa (partial sequence)
- B] Shine Dalgarno sequence upstream of TTG
start codon with possible signal peptide
evident at N-terminus.

ID-102

Clone RS-81

ATGACACAATCAGATGCATATCTCTCGTTAACCGCGAAGACACGCTTACA
GATCGCACAGGTAAATTATCATTTACTCGGATAAAGAGGGCTGTTAACAA
TATATGATAGAACATGTTAACCTAATACGATGGTGTTCACATCACTAATT
GAAAAGCTAGATTATTGGTTCTAATAACTACTATGAATCGGACCTTCTA
AAACAAATATAACCTTGAGTTATTGCCAAATTTGAGCATGCATACGCT
AAGAAATTGCTTTCTAAATTATGGGGCTTAAAATTATAATGCTT
ATGCTCTTAAT

MTQSDAYLSLNAKTRFRDRTGNYHFTSDKEAVEQYMIEHVEPNTMVFTSLIE
KLDYLVSNYYESDLLKQYNLEFICQIFEHYAKKFAFLNFMGALKFYNA
LN

Sequence description:

- A] Length: 318 bp - 106 aa (Partial sequence)
- B] Shine Dalgarno sequence present upstream of
ATG start codon, no obvious signal peptide

FIG. 1 CONT'D

ID-103

Clone 2-11A

ATGGTATTATGGCAAATAAGAAAAACAAAAGGAAAGAAAACCAGAA
GACCTACTAAGGCAGAAATAGAGCGTCAAAGAGCTATTCAAAGGATGATT
ACTGCTCTGTTAACATTATTCTCTTGGTATTATCAGATTAGGTA
TTTTGGTATTACAGTCTATAACGTCACTCGTTATGGTAGGTAGCTTGGC
TTACTTATTATTGCGGCAACTTAATCTACCTTATTTCTTAAATGGTTG
CGAAAGAAAGATAGCTTAGTAGCAGGTTTTGATAGCTTCTTAGGATTA
TTGATTGAGTGGCATGCTTACCTTCTCAATGCCTATTGAAAGATAAA
GAAATTTCGCTTCAACTGCTCGATTAATTGTGTCGATTAAATGCAATTAA
AAATCACTGTTTGCCGGTGGAGGTATGTTGGTGCCTTGAATTACAAGC
CAATTGCTTCTCTTCTAATATTGGTGCCTATATGATTGGTGTCTCTTC
ATCATTGGGTCTCTTTAAATGAGTTCTGGAAAGTTATGACATCGTCG
AATTATTAGAGCTTAAAGATAAGCGAGAGATGAAAAAGCAATCGCTGAACA
AAAAAGGAGCGTTGCTAACGAGAGATGAAAAAGCAATCGCTGAACA
AGAGCGCATAGAGCGTCAAAAGCTGAAGAAGAAGCTTATTAGCTTCGG
TTAATGTAGACCCCTGAAACGGGTGAGATTCTAGAGGATCAAGCTGAGGAC
AATTGGATGATGCGCTACCACCTGAGGTAGTGAACATCAACTCCGGT
ATTGAGCCAGAGATCCITGCTTATGAGACATCGCCTCAAAATGATCCTT
ACCAGTAGAGCCGACAATTATTAGAAGACTATGATTGCCGATTCTAA
TATGAGAGAAAATGATGAGGAAATGGTTATGATTAGATGATGATGTAG
ATGATAGTGTATAGAAAATGTCGACTTACACCTAAACGACACTGGTT
ATAAATTACCAACGATAGATTATTGACCCAGATAAGCCTAAAATCAAT
CCAAAGAAAAGGATTAGTCCGAAAGAATATCAGAGTTAGAAGAAACA
TTAGAAGTTGGTATCGATGTAAGAGTGAACGTGCTGAAATTGGACCA
TCAGTTACTAAATATGAAATTAAACCAGCAGTGGAGTTCGTGTGAATCGT
ATTICAAATCTATCTGACGACCTAGCTTCTGCTTGCAGAAAAGATGTG
CGTATAGAACCAATTCTGGAAAATCATTAAATAGGTATTGAAGTTCCT
AACTCAGAAATTGCAACGGTTCTTCCCGCAACTTGGGAACAATCTGAT
GCCAATCCTGAAAACCTTTAGAAGTACCAACTAGGAAAAGCTGTTAACGG
CAATGCTCGCAGTTAACCTAGCTAGAATGCCGATCTTGGTAGCTGG
TTCAACTGGTTAGGTAAATCTGTCAGTTAATGGAATTATTCAAGTAT
TTGATGAAGGCACGTCCAGATCAAGTTAAGTTATGATGATTGATCCAA
AATGGTTGAATTATCTGTTATAATGATATCCACATTATAATCCCTGTT
GTAACCAATCCCGTAAAGCAAGTAAGGCACTCCAAAAGTTGTTGATGA
AATGGAAAATCGATACGAGTTATTAGCAAAATTGGTAGCTGAGCAAAAC
CAGGTATAATACAAAGGTTGAAGAGTTAATGCTTCTGAGCAAAAC
AAATGCCCTTGCCTTAAATCGTTGTCATTGATGAAATTGGCTGACTTGAT
GATGGTTGCTAGTAAAGAAGTTGAAGAGATGCTATTATCGTTGGGGCAAA
AGCACGTGCTGCAGGTATCCATATGATTCTGCAACTCAACGTCCATCCGT

FIG. 1 CONT'D

AGATGTTATTCTGGTTGATTAAGCAAATGTTCCGTGCGTATTGCATT
 GCTGTTCAAGTGGTACTGATAGCCGTACGATCCTTGATGAAAATGGTGCT
 GAAAAGCTCTGGGACGGGGTACATGCTCTTAAGCCTATTGATGAGAAT
 CATCCAGTACGACTACAAGGTTCTTATTTCAGATGATGATGTTGAAAGG
 ATCGTTGGTTTATCAAAGACCAAGCCGAGGCTGACTATGATGATGCCCTT
 GATCCTGGAGAAGTATCTGAAACAGATAACGGCTCTGGTGGCGCG
 AGTACCTGAAAGTGATCCTCTTTGAAGAAGCCAAGGGACTCGTTTAGA
 GACGCAAAAGCAAGTGCCTCAATGATTCAACGCCGATTGCTGTTGGTT
 CAATAGAGCAACAAGACTAATGGAAGAATTAGAACGAGCAGCGGGGGTTATTG
 GTCCAGCAGAAGAACCAAGCCACGAAAAGTTTAATGACTCCAACCTCG
 AGTGAATAA

MVFMANKKTKGKKRRPTKAEIERQRMITALVLTIIFFGIIRLGIFGIT
 VYNVIRFMVGSLAYLFIAATLIYLYFFKWLRLKDSLVAGFLIASLGLLIEWHA
 YLF SMPILKDKEILRSTARLIVSDLMQFKITVFAAGGMLGALIYKPIAFLFSNIG
 AYMIGVLFIILGLFLMSSLEVYDIVEFIRAFKNKVAEKHEQNKKERFAKREMK
 KAIAEQUERIERQKAEEEAYLASVNVDPETGEILEDQAEDNLDDALPPEVSETST
 PVFEPEILAYETSPQNDPLPVEPTIYLEDYDSPIPNMRENDEEMVYDLDVVDD
 SDIENVDFPKTTLVYKLPTIDLAPDKPKNQSKEKDLVRKNIRVLEETFRSGI
 DVKVERAEIGPSVTKEYEIKPAVGVRVNRISNLSDDLALALAAKDVRIEAPIPGK
 SLIGIEVPNSEIATVSFRELWEQSDANPENLLEVPLGKAVNGNARSFNLARMPH
 LLVAGSTGSGKSVAVNGISSLMKARPQVKFMMIDPKMVELSYNDIPHLLI
 PVVTPRKRASKALQKVVDENMRYELFSKIGVRNIAGYNTKVEEFNASSEQK
 QMPLPLIVVIVDELADLMMVASKEVEDAIIRLGQKARAAGIHMILATQRPSVD
 VISGLIKANVPSRIAFAVSSGTDTSRTILDENGAEKLLGRGDMLFPIDENHPVRL
 QGSFISDDDVERIVGFIKDQAEADYDDAFDPGEVSETDNGSGGGGGVPESDPL
 FEEAKGLVLETQKASASMIQRRLSVGFRATRLMEELEAAGVIGPAEGTKPRK
 VLMTPTPSE*

Sequence description:

- A] Length: 2451 bp - 817 aa (Full-length gene)
- B] Shine Dalgarno sequence present upstream of ATG start codon, possesses a potential signal peptide

ID-104

Clone 2-18/22b

FIG. 1 CONT'D

ATGTCACAAGAGCAAGGAAAAATTATATTGTAGAAGATGATATGACGAT
TGTGTCACTTTAAAAGATCATTTATCAGCTAGCTATCATGTCTCTAGTGC
AGCAATTTCGTGATGTGAAACAAGAAAATTATCGCATTTCAACCCGATTG
ATACTAATGGATATTACGTTACCTATTAAATGGTTTACTGGACTGCAG
AATTGCGTAAGTTTAACAATTCTATTATTCATTTCATCTAGTAATGA
TGAAATGGATATGGTTATGGCATTAAATATGGGGGGTATGACTTTATTTC
AAAACCATTCTCTAGCTGTATTAGATGCTAAGCTAACTGCTATTAAAG
GAGAAGTCACAATTATCCAACAGGAATTAACTTTGGGGGATTACGTT
GACAAGAGAAGGGTTATTGTCTAGCCAAGATAAAAGAGGTTATTATCGC
CAACAGAAAATAAAATCCTATCTATCTGCTCATGCATCCTAACAGTAG
TCTCAAAAGAGTCTTATTAGAGAACTTGGAAAATGATAGTTTATTG
ATCAAAATACACTTAATGTTAATATGACACGCTTACGTAAAAAAATTGTCC
CAATAGGTTTGATTACATTACAGTGAGAGGAGTTGGGTATTACTAC
AATGA

MSQEQQKIVYIVEDDMTIVSLLKDHSASYHVSSVSNFRDVKQEIIAFQPDLILM
DITLPYFNGFYWTAEERKFLTIPIIFISSNDEMDMVMALNMGGDDFISKPFSLA
VLDAKLTAILRRSQQFIQQELTFGGFTLTREGLLSSQDKEVILSPENKILSILLM
HPKQVVSKESLLEKLWENDSFIDQNTLNVNMRKKIVPIGFDYIHTVRGVG
YLLQ*

Sequence description:

A] Length: 669 bp - 223 aa (full-length gene sequence)

B] Shine Dalgarno sequence present upstream of a GTG start codon. Was not identified directly by LEEP. This gene was found upstream of gene ID-10 described in WO 00/06736.

ID-105

Clone 2-20

ATGTATCAAACCTCAGACAAATAAGGAAAAATTGTTTATTGGAAATT
TTTATCCCAGTATTGATTATCAATTGCTAATTTCAGCTACTTTATTGA
TTCGGTTATGACTGGACAGTATAGTCAGCTACATTGGCAGGTGTGTCAAC
TGCTAGTAATTATGGACTCCGTTTCGCTTATTAGTAGGTATGATTCA
GCATTAGTACCACTAGTTGGTCAACATTGGGTAGAGGAAATAAAGAAC
AATTGCGCACAGAATTCTCATCAATTCTATATTAGGTTGATACTGCTTA
ATATTATTAAATCATGCAATTATTGCTCAACCTGCTTGGGGAGTTGG

FIG. 1 CONT'D

GTTAGAAGATGAAGITCTAGCAGTTGGCGTGGTTATTAAATTATATGT
 TGATTGGAATCATGCCGCTGGTGTGTTAGCATTGCCGTTCAATTCTTGA
 TGCATTGGGTTAACAGGTTATCTATGTATCTGATGCTTTAATTCTACCC
 TTTAATTCACTTTTAATTATATGCTTATCTACGGTAAATTGGTATGCCTA
 GACTAGGAGGTGCGGGGGCAGGTCTTGGAACTTCCTTAACCTATTGGGCTA
 TTTTATTGGTATTATTATGTGATGTCACCTCATCCTCAAATTAAAACATA
 TCATATATGGACTCTGGAAAGAACATTTAGCTCCTTGATTATTGAAGATAT
 TCGATTGGGATTACCGATTGGTTACAAATTTCAGAAGTTGCAATTTC
 GCAGTAGTAGGCTTATTCAATGGCAAAATTTCATCAATCATTATTGCAGCA
 CATCAGGCTGCTATGAATTTCATCATTAATGTATGCATTTCCTTAAGTA
 TTTCCACTGCTCTAGCTATTACAATATCGTTGAAGTAGGGCAGAGCGCT
 TTCAGGACGCAACCACTTATAGTAGGATAGGACGCTAACAGCGGTAGGG
 ATTACATCAGGAACCTTACTATTTCATTCGTGAGAATGTAGCAG
 CAATGTATAATAGGCCCTCACTTGTGCTTACAGCTCAATTCTAAC
 TTATAGCTCTTTCCAGTTGCAGATGCTTATGCAGCTCCTGTACAGGGG
 ATTTACGAGGCTATAAGGATACAACAAAACCATTATGATCGGTGCGGG
 CTCTTATTGGTTATGTGCTTGCCTATTGGATTGGTTAACACAGGTATTGGT
 CCAGTTAGGTCCGTTGCCTATTGGATTGGTTAACACAGGTATTGGT
 TGTGGTCTATTCTAAACCAACGTCTGAAAAGATTAAGAAGTTGTATTAT
 TAA

MYQTQTNKEKFVLFLKLFIPVLIYQFANFSATFIDSVMQYSQLHLAGVSTAS
 NLWTPFFALLVGMISALVPVGQHLGRGNKEQIRTEFHQFLYGLLILSILFLI
 MQFIAQPVLGSLGLEDEVLAVRGYNMLIGIMPLVLFICRSFFDALGLTRL
 SMYLMLLILPFNSFFNYMLIYGKFGMPRLGGAGAGLGLTSLYWAIFIGIIIVMS
 LHPQIKTYHIWTLERIKAPLIIEDIRLGLPIGLQIFAEVAIFAVVGLFMAKFSSIIIA
 AHQAAMNFSSLMYAFPLSISTALAITISFEVGAERFQDATTYSRIGRLTAVGITS
 GTLLFLFLFRENVAAMYNSAPHVAITAQFLTYSLFQFADAYAAPVQGILRG
 YKDTTKPFMIGAGSYWLICALPLAVILEKNSQLGPFAYWIGLITGIFVCGLFLNQ
 RLQKIKKLYY*

Sequence description:

- A] Length: 1341 bp - 447 aa (full length gene)
- B] Shine-Dalgarno sequence present upstream of ATG start codon, There is a potential signal peptide sequence

ID-106

FIG. 1 CONTD

Clone 2-4A

TTGCTAGTTCTCTAGTTCTTCATTTCATTTCTGTCAATTCTCGTCGTT
 GTCTTCATCAACACGAAATAAGTCTATAAACTTATCAAATAATTCTAGA
 CTTATTATATCAATTTCATAAAAATGCTATAATAAAAACCATGTCATTTC
 TTAAAAAATTAGAAATCCATACGGTGAACATAACCGTTAAAGAAACTCCTGA
 AGATTATTTTGATTCACCGTAAGATTAGACATTTCGCGTGTAAAAAA
 CATGTACTTATAAACAAATGAATTCTTAATTGGCAAACGTGCGTCCAAGAA
 AACGATACTTACCTTAATCTTGATGATGAGGATTACCTACTAAAAAA
 ATTCCCTCTGGCAGAGCAGAGCTTATTGATTGTTATGAGGATGAAACAT
 CTTATTATCGTTAATAAACCTGAAGGTATGAAAACTCACGGTAACCAACCA
 AATGAAATAGCACTGTTAAATCATGTATCTGCCTATTCTGGACAAACATGC
 TATGTTGTTCATCGCCTAGATATGGAGACCAGTGGAGCTGTTTATTGCT
 AAAAATCCATTATACCTCCCTTATCAATCAACGCTTAGAACGAAAAGAA
 ATTTGGCGTGAATATTGGGCTTATGTTGAAGGAAAATTTCACCTAACCAT
 CAAGTTTGAGAGACAAAATTGGACGGAACCGTCATGACAGACGTAACG
 AATCATTGATTCTAAAAACGGTCAACATGCTATGACAATCATTGACGTTT
 GAAGTATATCCAAAATAGTAGTCTCATAAAATGCCACTGGAAACCGGAA
 GAACCCATCAAATTGCACTTCACTTATCTCATCACGGACATCCTTAATAG
 GAGATCCCCTCTACAACCCCTTCTAATAATGAAAGGTTAATGCTACACG
 CTCACCGATTGACTCTATCCCATCCATTGCGAAACTATTAGCGTAG
 AGGCCCTTCATCTACTTTCGAGAAGGTTAAACAATTATAAAAAGGAG
 TTGGATAA

MLVSSLVSCSFFLVISSLSSSTRNKSINLSNNFIDLLYQFSIKCYNKTMFSLKIR
 NPYGEHTVKELLEDYFLIPRKIRHFLRVKKHVLINNEFINWQTVVQENDTITLIF
 DDEDYPTKKIPLGRAELIDCLYEDEHLIIVNKPEGMKTHGNQPNEIALLNHVS
 YSGQTCYVVHRLDMETSGAVLFAKNPFLINQRLERKEIWREYVALVEGKF
 SPKHQVLRDKIGRNRRKRIIDSNGQHAMTIIDVLKYIQNSSLIKCRLETG
 RTHQIRIHLSSHGHPLIGDPLYNPSSNNERMLHAHRLTLSHPLTCETISVEAPS
 STFEKVLNNYKKGVG*

Sequence description:

- A] Length: 1029 bp - 343 aa (Full length gene sequence)
- B] No obvious Shine-Dalgarno sequence upstream
 of the putative TTG start codon. Possesses a
 potential leader peptide sequence.

FIG. 1 CONT'D

ID-107

Clone 2-54

GAACTAAATGCAACTCAACCTAATAATAGAACTACCTATATTACCCGAA
 AGCAGTCATTCCATTGCAGAACACAGAGATTCTGTAGAATCAAAGGG
 TTCTCGGTTGCATTACTTAATAGCGATGAATTAGAAAGACAGCGGGAGA
 GGATAGAGGTTTGAAAGGGATAAGTTGAGGTCTTGGATATCATTCTAA
 GGGAGAGTTATCGACAAGTAATGTCATAGTAATACGGACATTGCTAGTC
 AGATATCGTGGGCTTAAAAAGAATGCGATGCAGGAACACCATCTTACT
 AAAACATTCTCTAAAAGGATGGAAAGTTATCGTCTGTTATAGAGGGGAT
 GCTTGCTATTGGCAAAGAGAAAGTAGAGAGAAAGAAATAATAGTGGTA
 ATTATGGCAAAATTAAAAGCTAAGGCACACTGCCCTGTTGCTGTGTTG
 ATAATTGAATTGAAAGATATAAAATCTTATTTCATATTATGTCACT
 AAACCATCAGCTCAAATTACCTAAAGGTGCTACTTCTGCTAAAACAGA
 AGTATATAGGGGAGGGAGATTGGGAGAAAAATAAAGATAATGTGTTG
 GTTACCGTATCCCCTCATTATTGAAAACCCAAAAAGGAACCTTACTTGC
 GAGCTGATGAAAGAATTGAGCAAGCTTGTGATTGGGAAACATAGGAATG
 GTTATTGCCGTAGTGAGGATGATGGTGTCACTGGGGAAAAAGAGAAC
 TATTGTCAATCTCCGTAAATAACCTAGAGTCCGCTAGTTACTAGTGGTGA
 CTATAGTGGCTCACCTATTAATATGGATATGGCATTAGITCAAGATACTAG
 CTCCAAGACGAAACGTATTTTCAATATATGATATGTTCCAGAAGGAAG
 AGGCATTAGTATTGCTAACACACCTGAAAAAGAATATAACCAAATCG
 GAGGACAGTCTTATCTAATTATATAATGGAAAGAAATCGAAGGTTT
 GAGGATGAGTGGTAGGACATGGTCATCACCTAGAGATATAACAGCAAGT
 CTTCGTCAGAAAGGCATGAAATTGGGAATAGGACCTGGAAAAGGTAT
 AGTTTAAATGGGGCCACCGCTGGTGTATTATTCTCGCTATTCT
 ACGAATTGAAATCTCATCTAACAGAGGTTACAATCTCACGCCATTAT
 TCAGACGACCATGGAAAACGTGGCATACTGGAAAAGCAGTTAATGATAA
 CCGTATACTTCTAACAGGAAATTCACTCCCTAACATGGATAATAA
 AAAAGAACAAAATACAGAACCGTACCCGTTAACATTGAAAATGGGACA
 TTAAGTTATTGAGGAATCTAACCTGGTAACCTAGAAGTAGGCCACAAGTA
 AAGACGGCGGGAGACTGGCAAAACCATGTTAACCGATATAAGGAAATT
 CATGATGCTTACGTCACACTCAGCTATTGCTTGGCATGACAAAAAA
 GAGTATTTATTAGTGAATGCTAACAGGAAAGAAGTGGCTAACACTCTGTT
 TGGATATGCACGCTAGCGCAAGTTAACAGGAAAGTGGCTAACACTCTGTT
 ATATCACCACATTCAAGATGGTGTCTAACACTCTGTTAACAC
 ACTAACATAATGATCAATTGGTGTCTTATGAACATAGAGAAAACATCA

FIG. 1 CONT'D

AAATAGTTTACTTAAATTACAAAGTTTAATTGGAGTTCTTAGTC
AATACAGAGAAGCAAGGCACTTATGGGAGAAAATGGCAGCAAATTGGCA
TGTTITGTTAAATTATTTATGA

ELNATQPNNRTTYIIPESSHSIAEQQRFLIESKGSSVALLNSDEFRKTAGEDRGF
ERDKLRSLDIIPKGDLSTNVIGNTDIASQISLGFKKNAMQEHHLTKTFSQKDG
KLSSVIEGMLAIGKEKVEKEIKYSGNLWQKLKAKAHCLVCCVDNLFEDIKS
YFQYYCHLNHQLKLPKGAILSAKTEVYRGGDFGRKNKDNVFGYRIPSLLKTQ
KGTLLAGADERIEQACDWGNIGMVIIRSEDDGVTWGKRETIVNLRNNPRVPL
VTSGDYSGPINMDMALVQDTSSKTKRIFSIYDMFPEGRGVISANTPEKEYTQI
GGQSYLNLYNNGKKSKVFTIRDKGIVYNFKGKKTDYHVITETTKSDHSNLGDI
YKGKQLLGNIYFTKHKTSPFRLAKSSYYVMSYSDDDGRTWSSPRDITASLRQ
KGMKFLGIGPGKGIVLKWGPHAGRIIPAYSTNWKSHLRGSQSSRLIYSDDHG
KTWHTGAVNDNRILSNGEKIHSLTMDNKEQNTESVPVQLKNGDIKLFMRN
LTGNLEVATSKDGGETWQNHVKRYKEIHDAVQLSAIRFEHDKEYILLVNA
NGPGKKCQDGYARLAQVNRRNGSFKWLYHHHIQDGSAFAYNSVQLNNDQFG
VLYEHREKHQNSFTLNYKVFNWSFLSQNTEKQGTLWEKMAANWHVLFKFYL
*

Sequence description:

- A] Length: 2052 bp - 684 aa (partial gene sequence)
- B] N-terminus has yet to be determined

ID-108

Clone 2-61

ATGCCTAAATTAAATCGTATCTTCCCTCTGCATTTATTATCCCTGACTTGTG
TAAACTCTGTGCAAGCTGAAGAACATAAAAGATATTATGCAAATTACCGA
GAAGCCGGATATGATGTTAAAGATAATAATAAAACCTAAAGCGTCTATCGTT
ATTGACAATAAAAGGTCATATTTGTGGGAAGATAACGCCGATTAGAACGT
GATCCCGCTAGCATGTCTAAAATGTTACTTATATTACTATTGAAGACT
TAGCTAAAGGAAAAACAAACCTCACACACCACAGTGACTGCAACAGAAACA
GACCAAGCCATAAGTAAGATTATGAAATTAGTAATAACAATATTATGCT
GGGGTTGCTTATCCTATTCTGAAC TGATTACTATGACGGCTGTCCCGTCA
TCTAATGTAGCAACTATTATGATTGCTAACCACTTATCACAAACAATCCT
GACGCCTTATTAAACGAATCAATGAAACCGCCAAGAAACTCGGTATGAC
AAAAACTCACTTTATAACCCCAGTGGGGCGGTAGCGAGTGCTTTAATGG
ACTTTACTCCCCAAAAGAATACGATAACAATGCTACTAACGTTACGACTGC

FIG. 1 CONT'D

SUBSTITUTE SHEET (RULE 26)

ACGTGATCTATCAATTAAACCTATCATTCCCTAAAAAATACCCGTATATA
 CTGAACATACAAAATATCCTGAAGTCAGGCCATGGTCGGAACTCCTTAT
 GAAGAAACATTACAACCTATAACTACTCTACCCCCGGCGCTAAATTGGA
 TTAGAAGGAGTAGATGGCTTAAAAACTGGTTCTAGCCCTAGCGCTGCTTT
 AATGCCTTAGTTACAGCTAACGCCAGAATACTCGCTGATAACTGTGGTT
 TTAGGAGTGGCGATTGGTCAGACCAAGACGGAGAGTACTATCGTCATCC
 GTTGTCACGCTTGTAGAAAAAGGTTAAAGACGCTAAAAATATTCT
 TTCTAAAACCTGTATTAAAAGCCCTAACCTAAAAAGAAGTTACTAA
 AACCAAAACTAAATCTATTCAAGAACAGCCTCAAACAAAAGAACAGTGGT
 GGACAAAAACAGATCAATTATCCAATCACATTGTATCTATTAAATTG
 TTCTGGGACCACATCGCTAGCCTTGTCTTAGCTGGATAGTATTACTTAT
 AAAGCGCTCTAGATAA

MPKLIVSFLCILLSLTCVNSVQAEHKDIMQITREAGYDVKDINKPKASIVDN
 KGHILWEDNADLERDPASMSKMFPLYLLFEDLAKGKTNLNTTVTATETDQAI
 SKIYEISNNNIHAGVAYPIRELITMTAVPSSNVATIMIANHLSQNNDPAFIKRINE
 TAKKLGMTKTHFYNPSCAVASAFNGLYSPKEYDNNATNVTTARDLSILTYHF
 LKKYPDILNYTKYPEVKAMVGTPYEETFTTNYSTPGAKFGLEGVDGLKTGS
 SPSAAFNALVTAKRQNTRLITVVLGVGDWSDQDGEYYRHPFVNALVEKGFK
 DAKNISSKTPVLKAVPKKEVTKTKTTSIQEQPQTKEQWWTKTDQFIQSHFVS
 ILIVLGTIASLCLLAGIVLLIKRSR*

Sequence description:

- A] Length: 1188 bp - 396 aa (full length gene)
- B] Shine Dalgarno sequence present upstream of ATG start codon, possesses a potential signal peptide

ID-109

Clone 45

ATGACTGAAAAATATTATAATTGGGCAACGCTTGGAACCGGGCGTTATTGCC
 AACGAATTAGCCCAAGCACTGGAACGACGTGGACAAAAATTATATTCTGT
 AGCTAATAGAACTTACGACAAAGGACTTGAATTGCTAACAAATATGGTA
 TCCAAAAAGTTATGATCACATAGATCAAGTATTGAAGACCCCTGAAGTGG
 ATATCATTATATCTACTCCCCACAATACTCACATCTCATTACGAAA

FIG. 1 CONT'D

GGCTTAGCAAATGTAAGCACGTTCTTGCAGAAAAACTATTACTTAAA
 TAGTACTGAGCTAAAGAACCATAGATTAGCCGAAACTAACCATGTTGT
 CTTAGCTGAAGCCAATGACTATTTCATATGCCAATTACCGCCAATTAAA
 AACATTAGTGTAGTGGAAAATTAGGACCCTTAAAGTATTCAATGA
 ATTCGGAAGTTATAAAGAATATGATATGACTAACCGTTTCAGTCGTG
 ACCTAGCAGGCGGTGCTTGCTGGACATTGGTGTATGCACTTCTTGTAT
 TCGCTGGTTATGTCAGAACGACCTCACAAACATTACCTCTCAAGTTACATT
 TGCACCAACAGGGTTGATGAACAAGTTGGTATCCTACTAACCAACCCAG
 CAAATGAGATGGCGACTGTCAGCCTAGTTACATGCAAAACAAACCTAAA
 CGAGCAACTATCGCTTACGATAAAAGGCTACATTGAACCTTGAATATCCG
 CGAGGACAAAAGGCAGTTATTACTTAACTGAGGATGGGATCAAGATAT
 TATCGAAGCTGGCAAAACTGAAAATGCTCTCCAATATGAGGTAGCTGATA
 TGGAAAGAACCAATTCAAGGAAAATCACCATGTACTAAACTATACC
 AAAGATGTTATGGATATCATGACACAGCTACGTCAAGAATGGGGATTAC
 CTACCCAGAAGAAGAAAAATGA

MTEKYYNWATLGTGVIANELAQALEARGQKLYSVANRTYDKGLEFANKYGI
 QKVYDHIDQVFEDPEVDIYISTPHNTHISFLRKALANGKHVLCEKSITLNSTEL
 KEAIDLAEVNHHVLAEMTIFHMPYRQLKTLVDSGKLGPLKMIQMNFGSYK
 EYDMTNRFFSRDLAGGALLDIGVYALSCIRWMSEAPHNITSQVTFAVTGVDE
 QVGILLTNPANEMATVSLSHAKQPKRATIAYDKGYIELFEYPRGQKAVITYT
 EDGHQDIIEAGKTENALQYEVADMEEAISGKTNHMYLNYTKDVMMDIMTQLR
 QEWGFTYPEEEK*

Sequence description:

- A] Length: 984 bp - 328 aa (full length gene)
- B] Shine Dalgarno sequence present upstream of ATG start codon, possesses a potential signal peptide

ID-110

Clone 2-2

GTGTATTCTCCTGTTAAATCTTCTAAAGGAAAGTGTATTGTTAAAAGT
 GATTTCTAAAGAGCTTCATAGAAAGGAGAGGAAATATTGTTT

MYSPVKSSKGKVILLKSDFLKSFIERGNICF

FIG. 1 CONT'D

33 / 110

Sequence description:

- A] Length: 96 bp - 32 aa (partial sequence)
- B] GTG start codon - no obvious Shine-Dalgarno sequence
- Possesses a potential signal peptide

ID-111**Clone 2-3**

AAATACTGTATCATTGCAACCTCAAATGCAGGTTTGGAAACGAAGCATTT
ACAGGTGACAGCGATAAAAGACTGAAAATTATGGAACGAATTCTCCATA
TTTCCGTCAGAATTCTAAATCGTTCAATGGTGTATTGAATTCTCTCAC
CTAACGAAAGATGACTTAAGCGAAATTGTAGATTGATGCTTGATGAAGTT
AACCAAACAATTGGCAAAAAAGGAATTGACCTTGTGGTAGATGAAAATGT
TAAATCACACTTAATTGAACGGGTATGACGAAGCAATGGGAGTACGTC
CATTGCCCGTGTCACTGAGCAAGAAATTGAGATCGCATCACAGACTACT
ATCTCGATCATACAGACGTAAACACCTAAAAGCTAATTGCAAGATGGCC
AAATCGTCATTCTGAAAGATAA

KYCIIATSNAGFGNEAFTGDSDKDLKIMERISPYFRPEFLNRNGVIEFSHLSKD
DLSEIVDMLMDEVNQTIGKKGIDLVDENVKSHLIELGYDEAMGVRPLRRVIE
QEIRDRITDYLDHTDVKHLKANLQDGQIVISER*

Sequence description:

- A] Length: 429 bp - 143 aa (partial sequence)
- B] N-terminus yet to be elucidated. This gene was not in frame with nuc

ID-112**Clone 2-5****FIG. 1** CONT'D

ATGTCAATGAATTTCATTTTACCAAAATTGGTCTTATTAAATTATG
 GTGTGATGGTAACCATTATGATTCAACATGTGTTTTTTGGAACTAT
 TATAGGCGTGTAAATTGCTTAGTAAAGCGTACTAATTACATTCTCACA
 ATATTAGCTAATTCTATGTATGGGTATTCGGGACACCGATGGTAGTT
 CAAATTATGATTGCTTCGCATGGATGCATTAAACAATTACCAACAATT
 AGCTTGGTGTAGATTAGATTACACGACTTTACCTGGTATCATTA
 TCATTCCCTAAATAGTGGTGCCTATATTGGAAATTGTACGTGCAGGG
 TTGAGGCTGTACCATCTGGACAAATAGAACAGCTTACTCGTTGGGATT
 GACCTAAAAACACTTCGCTATGTTATCTTACCCCAAGCTTAAAAATA
 TTTACCTGCTCTAGGGATGAATTATTACAATTATAAAGATAGTGCCT
 CCTCAAACATTGGTGTACGGAAATTATGGAACGGAGCACAACTAGTGT
 AACGGCTACTTACTCACCAGTTGCACCGTTATTATTGCAGCATTACTAT
 TTAATGTTGACAACGATTCTCTCAGCTTGTAAAACAAATGGAGAAATAT
 CTTGGAAAGGGTAAAAATAGATGGTGA

MSMNFSFLPQYWNSYFNYGVMVTIMISTCVVFFGTIIGVLIALVKRTNLHFLTIL
 ANFYVVVFRGTPMVQIMIAFWMHNNLPTISFGVLDLFTRLLPGIIIISLN
 GAYISEIVRAGIEAVPSGQIEAAASLGIRPKNTLRVILPQAFKNILPALGNEFITI
 IKDSALLQTIGVMELWNGAQSVVTATYSPVAPLLFAAFYYMLTILSALLKQ
 MEKYLGKGVKIDG*

Sequence description:

- A] Length: 699 bp - 233 aa (full length gene)
- B] Shine-Dalgarno sequence preceded the 'ATG' start codon. Possesses a potential leader peptide sequence.

ID-113

Clone 2-7

ATGAAAGACCTATTACGAAATAGTCTAGAGCAAAGTGGAAATTAAAGTT
 TCAAGATATGATTTCATATTCTGTAGCAGCTTATTGAGTGTAGTTATT
 TATGTTCTATGCTTACGCATAGTGGAACTGCCTATAGTAAAAAGTT
 AATGTTCTTAATGACATTGACGGTCTGACTGCAACAGTAATGACCGTT
 ATTGGTAATAATGTAGCCTTGTCTTGGGTATGGTCGGTGCCTTGTCAAGTT
 GTTCGTTTCTTGGACAGCCATAAAAGATTCAAGAGATAAGTTATTTT
 TGGACCATAGTTGTGGTATCTGTTGTGGTGCCTGACTATGTGGTAGCT

FIG. 1 _{CONT'D}

GCATTAGGAAGTAGCGTTATCTTATCTTATTATGGGTATGGGACGTGTT
 AAAAACGAGAATCGTATGTTATTGATTGTAAAGTGCATAGAACACTAGA
 AGTGATTAGAAGGAATTTCCTCCAATATTTGACGGAAAAGCTGTTCA
 GCGTGTAAAAATTCAACAACATAACTATTGAAATGATTTCGAAATCTC
 TAGAAAAGATTACGATAAGCAACTCCATGTAGATAATCAGTTAACTGAAA
 AAGTGTACCAATTGGGAAATATTGATTATTCACATTGTTAGCCAAAGCG
 ACGAAAATCAATGGGTAG

MKDLLRNSLEQSGNLSFQDMILHILVAALLSVVIYVSAYTHSGTAYSKKFNV
 SLMLTVLTATVMTVIGNNVALSLGMVGALSVVFRFTAIKDSRDTVYIFWTIV
 VGICCGVGDYVVVAALGSSVIFILLWVMGRVKNENRMLLIVKCDRTLEVLEGI
 FFQYFDGKAVQRVKNSTNTIEMIFEISRKDYDKQLHVNDNQLTEKVYQLGNID
 YFNIVSQSDEING*

Sequence description:

- A] Length: 678 bp - 226 aa (full-length gene)
- B] ATG start codon is preceded by a Shine-Dalgarno sequence-Possesses a potential leader peptide sequence

ID-114

Clone 2-8

AAAAATTCACTTACGATTCACTTACGACTATATACTCAGAAGTACCAAAC
 CTAATCCAAGGTTGAAAAAAAGAAAGAAGGAAGTCAGTATGACAAACTAT
 AAAAACAACTTAAAGATGAGGCTATACGTGTTGAAGAGACAAACAAAAGA
 ATCATTTACGATGTTGATATTGCCCTGTTTCAGCTGGTGGATCTATTCA
 GCAAAGTCGCTCCTATGCAGTAAAGTCTGGAGCAGTTGTTAGATAAAC
 ACGTCATATTTCGTCAGAATCCTGATGTTCCACTAGTTGTTCTGAAGTAA
 ATGCTCATGCCATGATTGGTCATAATGGTATCATAGCTTGTCCCAATTGTC
 TACTATTCAAATGATGATTGCTTATAGAGCCCATTGTCAAAAATGGGGAT
 AGAGCGTGTATAGTTCCACCTATCAAGCTGTTGGGTTCAAGGTGCACG
 TGCTGTTGAAGAAACTAAGGAACAGTTGAGACAAGTTT

KFILDSFYDYILRSTKPNPRFEKRKKEVSMTNYKNNFKDEAIRVEETKESFYD
 VDIALFSAGGSISAKFAPYAVKSGAVVVDNTSYFRQNPDVPLVVPEVNAHAMI
 GHNGIIACPNCSTIQMMIALEPIRKWGIERVIVSTYQAVSGSGARAVEETKEQ
 LRQV

FIG. 1 CONT'D

Sequence description:

- A] Length: 499 bp - 165 aa (partial sequence)
- B] N-terminus has yet to be determined

ID-115**Clone 2-9**

ATGACAAATGAATTGATAATGCAAGCTTGAGTGGTATTTACCTAGTGAT
GGGAATCACTGGAAGAAATTAGAGGAGTCTATATCAGACCTTAAAAACT
TGGAATTAGTAAAATCTGGTACCACCAAGCATTAAAGGAACTAGCAGTG
ATGATGTAGGATATGGTGTATGATCTCTTGATTAGGAGAATTGACC
AGAATGGAACAATTAGAACAAATATGGTAGGAAAGAAGAGTATCTAAA
GCTTATTAAAGTCGTTAAAGGCAAATGGCATTAAACCGTTGCAGATATCGT
TCTTAACCATAAAGCCAATGGTGATCATAAAGAAAAATTCAAGTCATCA
AAGTCAATCCTGAAAATCGTCAAGAACGATTAAGTGAACCCSTATGAGATT
GAAGGATGGACGGGATTGATTCCCAGGTAGACAGGGTGAGTACAATGA
TTT

MTNELIMQAFEWYLPSDLGNHWKKLEESISDLKKGISKIWLPPAFKGTSDDV
GYGVYDLDLGEFDQNGTIRTKYGRKEEYLKLKSLKANGIKPFADIVLNHKA
NGDHKEKFQVIKVNPENRQEALSEPYEIEGWTGFDFPGRQGEYNDF

Sequence description:

- A] Length: 456 bp - 152 aa (partial sequence)
- B] ATG start codon is preceded by a Shine-Dalgarno sequence, no leader peptide sequence.

ID-116**Clone 2-10****FIG. 1_{CONT'D}**

ATGGAGGTTCTTATGAAGAAAGTGTAGTAAGTAGTCAGCTTTGGTTAGGG
ATTACGATAACGTTACAACCAGTAGTTGAGGCTAAGGGGCCAAAAGTAGC
TTATACACAAGAGGGAATGACTGCTCTTCGGACACAAATAAGATAAAAG
TCACTACTATTCTATTGACGAGATTCAAAAAAGCTTAGAAGGTAAAGAAC
CGATTACTGTTAGTTGATATTGATGATACACTGCTTTCACTGTTGATTTCAT
TTTCAATATGGTAAAGAATATGTAACCTGGATCGTTGATTTCAT
AAACAAAAATTCTGGGATCTTGTGCAAAACGAGGAGATCAAGATTCCAT
TCCCAAAGAATATGCTAAAAAATTGCTATGCATCAAAACGAGGAG
ATAAAATTGTTTATAACAGGTAGGACAAGAGGGTCAATGTATAAGGAG
GGCGAGGTTGATAAAACAGCTAAAGCCTAGCTAAAGATTAAATTGTA
CCATCTGAT

MEVLMKKVLVSSLLVLGITITLQPVVEAKGPKVAYTQEGMTALSDTNKDKVT
TISIDEIQKSLEGKKPITVSFDIDDTLLFSSQYFQYGKEYVTPGSFDLHKQKF
DLVAKRGDQDSIPKEYAKKLIAMHQKRGDKIVFITGRTRGSMYKEGEVDKTA
KALAKDFKFVPSD

Sequence description:

- A] Length: 516 bp - 172 aa (partial sequence)
- B] ATG start codon is preceded by a Shine-Dalgarno sequence, Possesses a leader peptide sequence.

ID-117

Clone 2-17

ATGCTTAAAGATTATTACTGAAGATGGGAATTGACAAAGATTAGTCGT
CGTTTCGTTGGATGTTAGTGGTTATCTATTGCTTATTATTGTCAAGGATGT
GTTTGGGCCTCAAATTATGATTGAGGGGGTATCAACTCCGAATGTTCAAGC
GCTTCGAAGAATTGAGCTCTTGTACCATTTAATTCTTCTGTAGTT
AGATCAGCTAACTAGCTTAAAGAGATTCTTGGTTATTGGTCAAAATGT
AGTGAATATTACTGCTGTTCCCTCTCATTATAGGGTTACTATCCCTAAAG
CCAAGTTACGGAAATATAAAAGCGTTATTACTGCTTCTGATGTCTC
TTTCATAGAGTGTACTCAAGTTGTTAGATATTAAATAGATGCTAATCG
GGTTTTGAAATCGACGATCTATGGACAAATACCTTAGGCCTTCGC
CCTATGGAGTTATCGAAACATAAAAGGTTGGCTTCAACTATTAGAAAATG
A

FIG. 1 CONT'D

MLKRLFTEDGELETKISRRFVWMLVVIYCLIIVRMCFGQPQIMIEGVSTPNVQRFG
 RIVALLVPFNSFRSLDQLTSFKEILWVIGQNVVNILLFPLIIGLLSLKPSLRKYK
 SVILLAFLMSLFIECTQVVLDILIDANRVFEIDLWTNTLGGPFALWSYRNIKG
 WLLTIRK*

Sequence description:

- A] Length: 516 bp - 172 aa (full-length gene)
- B] ATG start codon is preceded by an Shine-Dalgarno sequence. Possesses a potential leader peptide sequence. C-terminus need further confirmation.

ID-118

Clone 3-3

ATGAAAAAGCTTACTTTATTGGGATTAGATGGGACATTAATAGATTG
 TATGTACCAATTATGGAAGCTCTGAAGAACCTATCGTCATTTGGCTTA
 ATATTTGATAAAGAATTAAATCCATGAATATATTTACAGGAATCAGTGGGG
 CAATTATTGTAACCTTCAAGAGGAAGAGCAAATACCTCATGAAAAACT
 GAAAGCATATTTACAAAAGAACAGAAAGTCGAGATTCTAAATACATT
 TAATGCCATATGCAAAAGAGATTAGAATGGACCAAAGAACAGATATT
 CCCAATTTATGTATACACATAAAGGAGCAAGTACGCATTCACTGTTGGAA
 ACCTTGCAGATCTCTCATTATTTGATGAAATTAACTGGTGTTCGGGAT
 TCGAGCGAAAACCACATCCACAAGGGATTAATTATTTAGTTAACGATATT
 CTTAGATAAAATCAATGACTTATTACATAGGAGATCGCCACTAGATTGG
 AGGTTGCTAAAATGCTGGTATAAAATCCATAAACTTAAGGTTAGAGAATT
 CCAAAGAAAACTATAATATTCAAGTCTAAAGATATAATACACTTGATT
 TCACTCGTTGGATTAA

MKKLTIFIWLDLDGLIDSYVPIMEALEETYRHGLIFDKELIHEYILQESVGQLL
 VNLSEEEQIPHEKLKAYFTKEQESRDSKIHLMPYAKEILEWTKEQDIPNFMYTH
 KGASTHSVLETLQISHYFDEILTGVSGFERKPHPQGINYLVKRYSLDKSMTYYI
 GDRPLDLEVAQNAGIKSINRLENSKENYNISSLKDIISLDFTRLD*

Sequence description:

- A] Length: 627 bp - 209 aa (Possible Full-length gene)

FIG. 1 CONT'D

B] ATG start codon is preceded by an possible Shine-Dalgarno sequence. No obvious leader peptide sequence.

ID-119

Clone 3-7

ATGGAAAAAGAAAAAAAATTAGGTCTTTACCACTAACAAATGCTTGTCAATT
 GGCTCTCTATCGGTGGCGGAATCTTGATTAATGCAAAATATGAGTTCC
 AGAGCCGGTTGGTACCAATGCTTATTGCTGGTAATTACTGCTATCGGG
 ATGGGAACCTTCGTTAAGTTTCAAAATTATCTGAAAAAAGGCCGGAC
 CTAACAGCTGGAATCTTAGTTACGCTAAAGAGGGGTTGGAAACTTATG
 GGATTTAACTCTGCATGGGTTATTGGTTATCAGCTTGGCTTGGAAATGTT
 GCCTACGCTGCACTCTTATTCAAGTTCACTCGGTATTCTTAAATTCTTG
 GTAATGGAAATAATATCATCTCAATTATTGGAGCAAGTATAGTTATTGGG
 TTGTCCATTCTTAATTAAAGAGGTGTTAATACAGCTGCATTATTAATAC
 CGTAGTTACCTTIGCAAAATTAGTACCTGTTATTATTCTTAATTTCAGCG
 TTATTAGCTTCAAATTAAACATTAGTCTGATATCTGGGAAATGGAT
 TACATCAATCAATTTCACCAAGTCATTCAACTATGAAAACCGCTGTT
 GGGTATTATGGTATTGAGGGCGCCGTGTCTCAGGTGCTGCTAAAAA
 AACACTCTGATATTGGTAAAGCAAGTATCCTAGCATTATTCACTATGATT
 CACTTATGTATTGATTCTGTTTATCACTGGTATCATGTCACGTCCAGA
 ACTTGCAAACCTAAAAACACCAGCTATGGCTTACGTTAGAAAAAGCTGT
 TGGTCACTGGGGTGTATCTTAGTTAACCTGGTATCATTCAGTATT
 GGCCTATTCTGCTTGGACTTTATTGCAGCAGAATTACCATATCAAGCT
 GCTAAAGAAGGTGCTTTCTAAATTGGCAAAAGAAAATAAAAACAA
 AGCTCCAATCAACTCACTCTTAGTCACTATCTTGTGTACAAGCATTCTA
 ATCACGTTCTTATTCACACAAAGCTTATCGTTGGTTCGCATTAGCAT
 CATCTGCTATCTTAATTCTTATGCTTTACAGCACTATATCAATTACAATT
 CACACTCCGTGAGGATAAGTCACCTCAGGACATCAAAGAATTAAATT
 TCGGTATCCTCGCTACAATCTATGCTTACCTTATCTACGCTGGTGT
 TGATTACTTACTTTGACAATGATTGCTTATACTCTAGGTATGATTCTCTAT
 ATTAAAATGAGAAAAGATGACAAGCTGGCGTAATCATGGTCATAGCTGT
 TTCCAGTGTGAAATTGTTATCC

MEKEKKLGLLPLTMLVIGSLIGGGIFDLMQNMSSRAGLVPMLIAWVITAIGMG
 TFVLSFQNLSEKRPDLTAGIFSYAKEFGFMGFNSAWGYWLSAWLGNVAY
 AALLFSSLGYFFKFFGNGNNIISIIGASIVIWVVFILRGVNTAAFINTVVTFAK
 LVPVIIFLISALLAFKFNIFSLDIWGNGLHQSiFNQNSTMKTAVWWFIGIEGAV

FIG. 1 CONT'D

VFSGRAKKHSDIGKASILALFTMISLYVLISVLSLGIMSRPELANLKTPAMAYV
 LEKAVGHWGAILVNLGVIISVFGAILAWTLFAAELPYQAAKEGAFPKFFAKEN
 KNKAPINSLLVTNLCVQAFLITFLFTQSAYRFGFALASSAILIPYAFTALYQLQF
 TLREDKSTPGHQKNLIIIGILATIYAVYLIYAGGFDYLLTMIAYTLGMILYIKMR
 KDDKLGVIMVIAVSSVKLLS

Sequence description:

- A] Length: 1356 bp - 452 aa (partial sequence)
- B] ATG start codon is preceded by an possible Shine-Dalgarno sequence. Possesses a potential leader peptide sequence.

ID-120

Clone 3-8

ATGAAATTGAAAAACGGCAGGTCTATTATGTTGCATAACATTTGCTATT
 TGCTATGCTATACAGGCTTATTGGGGAGCTGTTCTAATATTAACTACGC
 TTCATAAGGCAATATTCCCTTTTGATGGGAGCTGGATTGCCTATATTAT
 TAATATTGTAATGTCAGTCTATGAGCGATTATATATAAAGCTTTAAAGG
 ATCTAGACTATTAATGGCAATCAAGCGTAGTGTCTATGATTTATCCTAT
 GCAACTTTATTGGTTAATTGTCGGCTATTTCATTGTCAATTGTCATTCCAGATT
 TGATTCTAGTTGAGTTCTTATTGGTTATTGATAACGGGAGCACTTGCTAA
 ATTGGTTAATAATCTCAATGAAAATAACAAATTCTGAGGCTTAAATTAA
 TATGGGAACAGATAAAGACTTAGTTCTACTTTAAGTGGTTATGCCAGCA
 GATTTGAAGCAAGTTTATCTGGTTAACAAATTACTAACCTCAGTTCC
 TCTATTGCGGCAACACTCTGAATGTTTGTAGTTTATTTCATTCAATTAA
 CGTTTGGCAAACAAGGAGCAGTGGGACGTCAATTAAATTGTTAATTGA
 TACCTATTAGGTTCAACAGGCAAAACATTCCATTACGTTCGTCATATCCTT
 CATCAACGTTCCATTGGTTTGTAGCCAAACTTAAAGCTATGATT
 TAGGAAGTTGACGGTTATTGGTATGTTGATCTTCCAATTCTTATGCTTT
 AACAGTTGGGTTTAGTTGCTTTACAGCTCTAATACCGGTTGTGGGAGC
 CTACATTGGTGTACAATCGGTTCATCTTAATTGCTACTGAATCGCTTACT
 GAAGCATTCTGTTGTTCTTGTACCTTTACAACAATTGAGGGAA
 ATGTCATTATCCGAAAGTTGTCGGTGGATCGATTGGACTGCCCTATGT
 GGGTTTAATGGCTATTACTATCGGAGGTGCTTATGGGGATCTTAGGCA

FIG. 1 CONT'D

TGTTACTTGCTGTTCTGTCAGCTACTATCTATCAGATTGTAAAAGATCA
TATTATCAAGCGACAAACGCTTAGAAATCGTCACGAACCTATCGTTAA

MKFEKRQVYYVVITFAICYAIQAYWGAVSNILTLHKAIFPFLMGAGIAYIINI
VMSVYERLYIKLFKGSRLLMAIKRSVSMILSYATFIGLIVWLFSIVIPDLISLSS
LLVIDTGALAKLVNNLNENKQISEALNYMGTDKDVLVSTLSGYSQQILKQVLSV
LTNLLTSVSSIAATLLNVFVSFISIYVLANKEQLGRQFNLLIDTYLGSTGKTFH
YVRHILHQRFHGFFVSQTLEAMILGSLTIVGMLIFQFPYALTGVVLVAFTALIP
VVGAYIGVTIGFILIATESLEAFLFVLFILLQQFEGNVIVYPKVVGGSIGLPSM
WVLMAITIGGALWGILGMLLAVPVAATIYQIVKDHIKRQTLRNARTYR*

Sequence description:

- A] Length: 1134 bp - 378 aa (full-length gene)
- B] ATG start codon is preceded by an typical Shine-Dalgarno sequence. Possesses a potential leader peptide sequence.

ID-121

Identical to ID-68, as described in WO 00/06736

ID-122

Clone 3-16

GTGATTACAATTAAAAAGGAATCTGTTATCAAACATTGAAGTATGCTTT
GGCATTATAATGGGATTATTATCTAGCTATTGTAATAGGTGGGCTCCTA
TTTGCATACTACGTTAGTCGTTCTCCGAAATTAAACCGATCAAGCTTAAAA
TCCGTTAACCTAGTTGGTTATGATGGTAATAATAAACTTATTGCCGATT
TAGGCTCAGAAAAGCGTAAAGTGTAGTGCAGGATAGCATTCACAAAT
TTGGTTAACGCTATCACTTCTATAGAAGATAAACGTTCTTAAACATAGA
GGTGTGATATTATCGTATTTAGGTGCAGCTGGCATAACCTTGTAGTA
GTAATACGCAAGGTGGTTCAACCCCTGATCAACAGTTGATTAAACTGGCTT
ACTTTCTACCAATAAAATCTGACCAACGTTAAAACGTAATCACAGGAA
GTTGGCTTGCCTCAAATGGAGCGTAAATACACCAAAGAAGAAATTCTT
ACTTCTATATTAATAAAAGTTATATGGAAATTATGGTATGAGA

FIG. 1 CONT'D

ACAAACAGCTAAATCATACTTGGTAAAGACCTAAAGGAATTATCTATTGCA
 CAACTTGCTTGCTCGCTGGTATTCCCTCAAGCACCTACACAATATGACCCTT
 ATAAAAAACCCAGAATCTGCTCAAACAAAGACGTAATACCGTTCTCAGCAG
 ATGTATCAAGATAAAAACATTCTAAAAGGAATACGCCAAGCTGTGC
 AACTCCAGTAACGTGATGGCTAAAAGAATTAAAGCAAAATCTACTTATCC
 AAAATATGGATAACTACTTAAACAAAGTTATTAGTGAAGTTAACAAA
 AAACTGGTAAAGATATCTTACTGCTGGCTAAAAGTGTATACTAATATCA
 AACTGATGCACAAAACAACATATGACATCTACAAACAGTGATACTTAC
 ATCGCTTATCCAACAAATGAATTACAAATAGCATCTACCATCATGGATGCG
 ACTAATGGTAAAGTCATTGCACAATTAGGCAGGCGTCATCAGAATGAAA
 TATTTCATTGGGACAAATCAATCTGCTTAAACAGACCGCGATTGGGGTC
 TACAATGAAACCTATCTCAGCTTATGCACCTGCTATTGATAGTGGTGTCA
 TAATTCAACAGGTCAATCATTAAACGACTCAGTTACTACTGGCCTGGTAC
 TTCTACTCAACTATGACTGGATCGTCAATATATGGGTTGGATGAGTAT
 GCAGACCGCTATTCAACAAATCACGTAACGTCCCTGCTGTCAAGAGCACTTGA
 AGCCGCTGGATTAGACGAAGCAAAATCTTCCTGAAAAATTAGGCATAT
 ACTATCCAGAAATG

MITIKKESVIKLKYAFGIIMGFIILAIIVIGGLFAYYVSRSPLTDQALKSVNSS
 LVYDGNNKLIADLGSEKRESVSADSIPLNLVNAITSIEDKRFFKHRGVDIYRILG
 AAWHNLVSSNTQGGSTLDQQQLKLA^YFSTNKSDQTLKRKSQEVLALQMER
 KYTKEEILTFYINKVYMGNGNYGMRTTAKSYFGKDLKELSIAQLALLAGIPQA
 PTQYDPYKNPESAQTRRNTVLQQMYQDKNISKKEYDQA^VATPVTDGLKELK
 QKSTYPKYMNDYLKQVISEVKQKTGKDIFTAGLKVTNINTDAQKQLYDIYN
 SDTYIAYPNNELQIASTIMDATNGKVIAQLGGRHQENISFGTNQSVLDRDW
 GSTMKPISAYAPAIDSGVYNSTGQSLNDSVYYWPGTSTQLYDWDRQYMGWM
 SMQTAIQQSRSRNVPAVRALEAAGLDEAKSFLEKLGIYYPEM

Sequence description:

- A] Length: 1386 bp - 462 aa (partial sequence)
- B] GTG start codon is preceded by an typical Shine-Dalgarno sequence. Possesses a potential leader peptide sequence.

ID-123

Clone 3-17

FIG. 1 CONT'D

43 / 110

ATGGCTAATGTATATGATTAGCAAATGAATTAGAACGTGCTGTTCGTGCT
 TTACCAAGAACGAGCTTAACTGCAAAGCAGCTATTGAAAATGA
 TGCAGATGCACAAGTGCCTGGCAAGACTTTGGCTACCCAATCAAAAGT
 TCAAGAAATGATGCAATCTGCCAAATGCCAAGTCAGAAGAACAGATG
 AAATGTCTAAACTGGGGAAAAAATTGAATCCAATGACCTTTAAAAGTT
 ATTTGACCAACAACACGGTTGTCTATATGTCTGATATCGAAAAAA
 TTGCTTGACCCATGCAGGACTTGATGTAA

MANVYDLANELERAVRALPEYQAVLTAKAAIENDADAQVLWQDFLATQSK
 VQEMMQSGQMPSQEEQDEMSKLGEKIESNDLLKVYFDQQQRSLSVYMSDIEKI
VFAPMQDLM*

Sequence description:

- A] Length: 336 bp - 112 aa (full length sequence)
- B] ATG start codon is preceded by an typical Shine-Dalgarno sequence. No obvious potential leader peptide sequence.

ID-124

Clone 3-26

ATGGCAGAAATCACAGCTAAACTGTAAAAGAATTGCGTGAAAAATCAGG
 TGCAAGCGTTATGGACGCTAAAAAGCATTAGTAGAAACTGATGGTGACC
 TTGATAAAGCGATTGAATTACTCGCAGGGTATGGCTAAAGCAGCT
 AAAAGCAGACCGTGTGCTGAAGGTTAACAGGGTGTATGTTGAT
 GGTAACGTTGCAGCAGTTATTGAAGTTAA

MAEITAKLVKELREKSGAGVMDAKKALVETDGDLDAKIELLREKGMAAK
 KADRVAAEGLTVYVDGNVAAVIEV

Sequence description:

- A] Length: 230 bp - 76 aa (partial sequence)
- B] ATG start codon is preceded by an typical Shine-Dalgarno sequence. No obvious potential leader peptide sequence.

FIG. 1 CONT'D

ID-125

Clone 3-33

ATGATAAAAAACCTGTTATTAAACAGGTTTATCATTAAATGACGGAAAA
CTGGACACAAATTATTTCTGTATAATTAAATATTATTCCTATCAGG
AGGTTATGATGACATTAGAGAACGATTAA

MIKNLLTGFLSFNDGKLDNYFSCIKYIISYQEVMMTLEKRF

Sequence description:

- A] Length: 134 bp - 44 aa (partial sequence)
- B] ATG start codon is preceded by an typical Shine-Dalgarno sequence. Possible potential leader peptide sequence.

ID-126

Clone 3-41

ATGAAAATAATAAAATAATGGTTTCTGAAAATTCTTATTACATA
TTATTGATTATTGCGGTTATTACAACCTTCAACTATTAA

MKNNKNNGFLKNSFIYILLIAVITTFQYYL

Sequence description:

- A] Length: 94 bp - 31 aa (partial sequence)
- B] ATG start codon is preceded by a possible Shine-Dalgarno sequence. Potential leader peptide sequence.

FIG. 1 CONT'D

ID-127

Clone 3-42

ATGTTAGATATTATCTTATCCGAATTCGCAAGGATTACTTGGTCAATT
TGGCAATTGGCGTGTTATCACTTTCTGTACTTAGACATAGCCGATCTCTC
TGCAGAAGGGGCTTCCCTATGGGGCTGCAGTTGCGCCTTATGTATCGT
TAA

MLDIILSGISQGLLWSIMAIGVFITFRILDIADLSAEGAFPMGAACALCIV

Sequence description:

- A] Length: 158 bp - 52 aa (partial sequence)
- B] ATG start codon is preceded by a possible Shine-Dalgarno sequence. Potential leader peptide sequence.

ID-128

Clone 3-43

ATGGAAATGCCTAAAAGAAATGAATTACTCAATAAAGAAATTAAAATGAG
TATTGATAAACTTAGATATAAGAACCGAGAGAGTGAACATGACAAGCGAC
CTACTTTTATTGGTAGTACTTATACTGTTACTGTAGCAGTTATATTGTC
GTTATTTAA

MEMPKRNELLNKEIKMSIDKRYKEPESEHDKRPTFYLVVLILVTAVILSLF

Sequence description:

- A] Length: 161 bp - 53 aa (full-length gene)
- B] ATG start codon is preceded by a possible Shine-Dalgarno sequence. Potential

FIG. 1 CONT'D

SUBSTITUTE SHEET (RULE 26)

46 / 110

leader peptide sequence.

ID-129

Clone 3-44

GTGGTAAGTAAATTGAGTTAACAAACGATTTGCATTGCTATTCATCA
 ATGCTAATTACGCAACACCTCTTACAGTATTGGGGAACCTTC
 TCTGAACGTGGTGGTATCGTCAACGTTAGAAGGAATTATGGTAATT
 GGAGCTTCTCAGGCGTTGTATTAA

MVSKLSLTIFALLFSSMLIYATPLIFTSIGGTFSERGGIVNVLEGIMVIGAFSG
 VVF

Sequence description:

- A] Length: 179 bp - 59 aa (partial sequence)
- B] GTG start codon is preceded by a possible Shine-Dalgarno sequence. Potential leader peptide sequence.

ID-130

Clone 3-46/47

ATGAGAATTATTGCAATAACTGAAAAGGTTATAAAAGAACTGTTCTGTGAT
 AAAAGAACACTTGCTATGATGTTTAGCACCTATTTAATTATGTTTGA
 TGAATGTTATGTTCTCGAATAGTAATACAAAAGTTAAGATTGGAACTA
 TTAACGTTAACACGAAGGTGTTCAAATTAGATAATTTAACCATATT
 AAGTGAGATCATTAAATTAACTCATCTGCTAAAAAGCACTCAAATCAA
 ATAAAATTGATGCTTATTCCGGAGGACAATAAATCTTATACTGTCTCT
 ATGCGAATACAGATTCTCAAAGACGACTTAAACAAGACAAGCTTTAAA
 ACCGCTGTTAACATGAACAGTAAGGAACGTTAAGGAACTGATTGCAAGTTAAAATT
 TTAGCTAATAAGAATCCGAAACTAGCACAATCCTTACAAACTCGCTCCAAA
 TATATCAAAGAAAAATATAATTACGGAAATAAAATACAGGCTTTTGC
 AAAAATGATAACCAACTAATGGGATTATGGTCTTCTGGTTTT

FIG. 1 CONT'D

47 / 110

MRJIIAITEKVIKELFRDKRTLAMMFLAPILIMFLMNVMFSANSNTKVKGITINV
NTKVVSNLNDNIKHIQVRSFKFNSSAKKALKSNKIDALISEDNKSYTIFYANTDS
SKTTLTRQAFKTAVNTMNSKELISQVKILANKNPKLAQSLQTRSKEYIKEKYNY
GNKNTGFFAKMIPILMGFMVFFLVF

Sequence description:

- A] Length: 558 bp - 186 aa (partial sequence)
- B] ATG start codon is preceded by a possible Shine-Dalgarno sequence. Potential leader peptide sequence. C-terminus has yet to be determined.

ID-131

Clone 3-48

GTGATTATCGTTATGAGTAAACATCAAGAAATTGGAGTACCTAGAAAAT
TTAGCTGTTGGTAAGAGGGTTAGTGTACGCAGTATTCAAATCATTAA

MIIVMSKHQEILEYLENLAVGKRVSVRSISNHL

Sequence description:

- A] Length: 100 bp - 33 aa (partial sequence)
- B] GTG start codon is not preceded by a obvious Shine-Dalgarno sequence. No obvious leader peptide sequence.

ID-132

Clone 2-c53

FIG. 1 CONT'D

SUBSTITUTE SHEET (RULE 26)

ATGTATAGAGAAATTACCGCTGTCGAACACGATCGCTTGTGAGCGAATCC
 AACCAAACAAACCTACTCAATCTCTTAATTGGCCCAAAGTAAAAGACAA
 CTGGGGTAGTCATTACTTGGCTTTTGACGGTAAACCCAAATTGCCAG
 CGCTAGTATTCTCATCAAATCACTCCTCTGGCTTCCATGCTGTATATT
 CCGCGTGGACCAATCATGGATTACTCCAATCTAGATATTGTAACTAAGGTC
 CTTAAGGACCTAAAGCTTGGAAAAAACAAAGAGCTCTTATCAAG
 TGTGATCCTCTCATCTATT

MYREITAVEHDRFVSESNQTNLQLQSLNWPKVKDWNWSQLLFFDGETQIASA
 SILIKSLPLGFMSLYIPRGPIMDYSNL DIVTKVLKDLKAFGKKQRALFIKCDPLI
 Y

Sequence description:

- A] Length: 326 bp - 108 aa (partial sequence)
- B] ATG start codon is preceded by an obvious Shine-Dalgarno sequence. No obvious leader peptide sequence.

ID-133

Clone 2-c59

ATGGACAAGAAAAAAATCTTAGTAACGGGTATTGTGCCTAAAGAAGGTCT
 AAGAAAGCTTATGGACCGATTGATGTTACTTATTCAAGATGCCATT
 TTCACGTGACTATGTGTTAGAGCATTTATCTGAATATGACGGATGGTTACT
 CATGGGACAAAAGGTGATAAAGAGATGATTGATGCAGGTGAAAACCTAC
 AAATTATTCCTT

MDKKKILVTGIVPKEGLRKLMDRFDVTYSEDRPFSRDYVLEHLSEYDGWLLM
 GQKGDKEMIDAGENLQIIS

Sequence description:

- A] Length: 215 bp - 71 aa (partial sequence)

FIG. 1 CONT'D

SUBSTITUTE SHEET (RULE 26)

B] ATG start codon is preceded by an obvious Shine-Dalgarno sequence. No obvious leader peptide sequence.

ID-134

Clone 2-c62

ATTCGAAAGATGACTACCAAAATTAGTTGGACAGGATCCAGAAGTT
 GTTGATTATGCTGGTCTGTTGAAAAACGCCGTCCAGTTAGAAAAAGCA
 GTTAAAATTCTTGCAAGAACAGAGAGCTACGAGAATGCTATCTGATTTC
 TTGCAAGAACAGAAAAATGGTAAGTGAATTGCTGAATTATGGCGATCAA
 AGAACATTGGTAATAAGGCCTCAAGAACAGGATGACAAGGCTATTA
 TACGCCCGAAGAACAGCCTAGCAGGATATCGTCAAAAGCTTAGTGAA
 GTGATAAAATATCATGAAGTAACGCAATATTCTTTACAAACAATGGTTT
 GAGTAAAGAATATGCTAATGATAAAGGGATTCAAATTATCGGTGATAT
 GCCAATCTACGTTCTGCCGATAGTGTAGAAGTTGGACAATGCCTGAAC
 GTTT

ISKDDYQNISFGQDPVVVDYAGLFKRRPVLEKAVKNFLQEERATRMLSDLQ
 EEKWVTDFAEFMAIKEHFGNKALQEWWDDKAIRREEALAGYRQLSEVIKY
 HEVTQYFFYKQWFELKEYANDKGIIQIGDMPIYVSADSVEVWTMPELF

A] Length: 459 bp - 153 aa (partial sequence)
 B] More sequencing is required to determine the
 N- and C-termini
 enzyme). - *Streptococcus pneumoniae* (63%)

ID-135

Identical to ID-108 described in WO 00/06736

Clone 2-c63

ID-136

Clone 2-c66

FIG. 1 CONT'D

ATGGCAAAACAGAAAAATAACTGGCGCCGTGTTGGAGTTGGTGTCTTAC
 ACTTGCTTCAGTTGCGACTCTGCTGCATGTGGAAGTAAATCAGCTTCCA
 GGATTCTAATGGAGCGATTAATTGGGCTATTCCAACAGAAATCAATACACT
 AGATTATCTAAAGTTACAGACACTTACTCAAATCTAGCTATTGGTAAC
 TAGTAGTAATTCCCTCGCTAGATAAAGATGGAAAGACAAGACCAGACTT
 GGCTACTAAAGTTGATGTTCAAAAGATGGCTTAACCTATACAGCTACATT
 ACGTAAAGGCTTGAAGTGGTCAGATGGCAGTAAACTACTGCAAAGGATT
 TTGTTTATTCACTGGCAACCTTACTGATCCTAAAACAGCTTCACAATATG
 CTTACCTGCTGTTGAAGGGCATGTGCTTAATGCCGATAAAATCACGAAG
 GACAAGAGAAAGACTTGAATAAGCTAGGTGTTAAGGCAGAAGGCATGA
 CAAAGTTGTTATTACTTATCTAGTCCGTCCGCAATTCTACTACCTT
 GCATTCACTAACTCATGCCACAAAAACAAGAAGTTGTTGAAAAATATGG
 AAAAGATTACGCAACTACTCAAAAAATACAGTTACTCAGGACCATA
 CTGTTGAAGGTTGGAATGGTCAATGGTACTTACGCTGAAGAAAAAC
 AAAAATTATTGGGACGCTAAAATGTAACAAAGAAGTTCGCATCCA
 GACTGTTAAAAAACAGATACCGCCGTTCAAATGTATAAACGTGGTGAGT
 TAGATGCAGCTAATATCTCAAATACTCTGCTATTATCAAGCTAATAAAA
 ATAATAAAAGATGTCACAGATGTTAGAACGCGACACTGCCTATAATGGAA
 TATAATACTACTGGTCTGTGAAAGGGCTTGATAATGTTAAGATTGTCGC
 GCCTTAAACTTAGCAACTAACGTAAGGAGTTGTTCAAGCAGCCGTTGAT
 ACAGGCTCAAAACCGGCAATTGCTTTGCACCTACTGGTTAGCCAAAACA
 CCAGATGGAACTGATTGGCAAAATATGTTGCCAGGTTATGAATATAAT
 AAAACTGAAGCAGCAAAACTCTTAGACTA

MAKQKNNWRRVGVGVLTLASVATLAACGSKSASQDSNGAINWAIPEINTLD
 LSKVTDYNSNLALIGNSSNFLRLDKDGKTRPDLATKVDVSKDGLTYTLRG
 LKWSDGSKLTAKDFVYSWQLVDPKTASQYAYLAVEGHVLNADKINEQEK
 DLNLGVKAEGDDKVVITLSSPSPQFIYLAFTNFMPQKQEVEVKYGKDYAT
 TSKNTVYSGPYTVEGWNGSNGTFLKKKNYWDANKVTKEVRIQTVKKPD
 TAVQMYKRGELDAANISNTSAIQANKNNKDVTDVLEATTAYMEYNTTGSV
 KGLDNVKIRRALNLATNRKGVVQAAVDTGSKPAIAFAPTGLAKTPDGTDLAK
 YVAPGYEYNKTEAAKLFR

Sequence description:

- A] Length: 1143 bp - 381 aa (partial sequence)
- B] Shine-Dalgarno sequence precedes ATG codon.
- Possesses a potential leader peptide sequence.

FIG. 1 CONT'D

ID-137

Clone 2-c67

TTGAGAGTTATGAAAATAAAGAAGAGTTGAAAAAAGAAATAAGTAAAAC
ATTGAGAAATACATTATGGAATTAAATAA
TATTCCAGAGAATCTAAAGATAAAAGAATTGATGAAGTTGATAGAACTC
CAGCAGAAAACCTTCTTCAGGTGGCT
GGACCAACTGGTCTTAATGGGAAGAAGATGAAAGAAAGGGACTCAA
GTAAAAACACCATCGGATAAATT

MRVYENKEELKKEISKTFEKYIMEFNNIPENLKDKRIDEVDRTPAENLSYQVG
WTNLVLKWEEDERKGLQVKTPSDKF

Sequence description

- A] Length: 234 bp - 78 aa (partial sequence)
- B] TTG start codon is preceded by a potential Shine-Dalgarno sequence. No obvious leader peptide sequence.

ID-138

Clone 2-c70

ATGTCAAAGTTGATAGTCAGAAAATAATTACTCCGATTATGAAGTTGTC
AATATGCGAGGGATTATTGCACTCAAAGATGGCATGCTAGCAATTACCA
CTAACAGTTGGAGTCTCTTTAATATTAGGGCAGCTTCCATIT
MSKFDSQKIIPIMKFVNMRGIALKDGMLAILPLTVVGSFLILGQLPF

Sequence description

- A] Length: 150 bp - 50 aa (partial sequence)
- B] ATG start codon is preceded by a potential Shine-Dalgarno sequence. Possesses a potential

FIG. 1 CONT'D

SUBSTITUTE SHEET (RULE 26)

leader peptide sequence.

ID-139

Clone 2-c71

GAGACCACCTCATCAGTTAAACCAGCAGGAATTGACCGTATCAATCATACC
 TCAACACCCCCGAAGAAAACTACCCCCAACATTGCAACGACGCATAGCTT
 CAAAGATCGTTGTGATACTTAGAAAGAATTACAATGAAGACATTGATGT
 TTGTTCTGGATTCAATTGTGGTATGGGAGAGAGCGATGAGGGGCTCATCAC
 ATTAGCTTCAGACTAAAAGAACTGAACCCCTATTCTATCCCTGTCAATT
 TTACTTGCTGTTGAAGGAACACCTCTGGAAAATATAACTATTGACTCCC
 ATTAATGCTTAAAAATTATGCCATGTTGCGTTTGTGTTTCCCTTCAAGG
 AATTAAGATTAAGTGTGGACGGGAGGTCCATTGAGAATTGAAATCAT
 TAGTCACCTTACTTGTTGACTCAACTTTTGGAAATTACCTAACAGAGG
 GGGTCGCAATCACATACCGATATTGAATTCTGGAAAAATTACAAC
 AATCATACTAAAAGGAATTAATT
 ETTSSVKPAGIDRINHTSTPPKTTPNIAHHFSFKDRCDTLERIHNDIDVCSGFI
 CGMGESDEGLITLAFLKELNPYSIPVNFLAVEGTPLGKYNLTPIKCLKIMA
 MLRFVFPFKELRSLSGREVHFENFESLVTLVDSTFLGNYLTEGGRNQHTDIEF
 LEKLQLNHTKKELI

Sequence description:

A] Length: 535 bp - 178 aa (partial sequence)
 B] N- and C-termini require verification

ID-140

Clone 2-c73

ATGCCGGTTGGACTGCACAGTCTATTCAAAGGCATTAGAAAAGCAT
 AATACTAAGGAAGGCACCTGGGAAAACAACCATTCTAAGTGGTTCTTA
 GTATTTACCAAGTTATCTCCTGATGGAGAGGAAATCTCGCGCATATT

FIG. 1 CONT'D

GATGCTAGTAGTGATATTCCCTTTGTTGATCCACAAGTCTGGCATAAAGTT
 TCGCCGAATAGTCCAGACTTAAGTTGCTATCTAACTTTACTGCCAAAAA
 GAAGATTACTCCATAAAAATATGGTCTACGCGCACACATTGAGGTT
 ATGCCAGTGCACCTCTTTATCTGAGAAGAGTAATATATTAGACCTGGG
 TGTGGTCAAGGGCGAAACTCACTTTATTATCGCTGCTGGGACATCAAGTG
 ACTTCTGTGATTCAAACGGACAGAGCCTGTAGCTTAGAAAATATGGCA
 TTAGAAGAAGAGCTCCTTACAATATAAAAAGGTATGATATTAACTACT
 GCTATTGAAGGGCACTATGATTATTATCAACTGTGGTATTATGTTT
 T

MPVWTAQSIKAFLEKHNTKEGTWAKLTILSGSLVFYQLSPDGEISRHIFDAS
 SDIPFVDPQVWHKVSPNPDLSCLTYCQKEDYFHKKYGLTRHSEVIASAP
 LLSEKSNILDLCGQGRNSLYLSLLGHQVTSVDSNGQSLVALENMALEEELPY
 NIKRYDINTTAIEGHYDFILSTVVFMF

Sequence description:

- A] Length: 563 bp - 187 aa (partial sequence)
- B] N- and C-termini require verification

ID-141

Clone 2c76

ATGACAAAGCAAATAATTGCCATTGGGCTGAAGATGAAGACCATTGAT
 TGGAGTTAATGGCGTTTACCATGGAGGCTTCCTAAAGAGTTACATCACTT
 CAAAGAAACGACCATGGGGCAGGCTTGCTTATGGGACGAAAGACCTTG
 ATGGAATGAACCGTCGTGTTTACCTGGTAGAGAGAGACAATCATCTAACAA
 AAGATGAACAATTCCAAGCAGATGGAGTGACAGTCCTAAATAGTGTGAA
 CAAGTTATAAAATGGTTCAAGAACATAATAAGACCTTATTATTGTAGGT
 GGTGCAAGTATTATAAGCATTTCTGCCTTATTGTGAAGCAATCATAAAA
 ACTAAAGTTCATGGAAAATTCAAAGGTGATAACCTATTTCCTGATGTTAAT
 CTATCTGAGTTT

MTKQIIAIWAEDEDHLIGVNGGLPWRLPKELHHFKETTMGQALLMGRKTFDG
 MNRRVLPGRETIILTDEQFQADGVTVLNSVEQVIKWFQEHNKTLFIVGGASI
 YKAFLPYCEAIKTKVHGKFKDNYFPDVNLSEF

FIG. 1 CONT'D

Sequence description:

- A] Length: 417 bp - 139 aa (partial sequence)
- B] ATG start codon is preceded by a Shine-Dalgarno sequence. No leader peptide sequence

ID-142**Clone 2-c78**

TTGTGGCCAAACTGTCCCCGCTTATTAATAGCACTTGTTCACCATTGAA
GATATCTAACATCAGGTGCTCATAGCAACCTATTATAATGGGGGTATA
CTTGGCGGGACAATTGAGTAGTGGCGACAGCACCACTTCTTCTATGGCA
TTGACAGCTATGCTAGGATTAACCGGAATGCCTATGGCTATAGGAGCCTTG
TCTGTCTTGGTTCGTCTTTAATGAATGGTGTACTTTCCATAAAATTAAAAC
TTGGAAGTCGTAAGATAATATAGCTTGTGCTGTTGAGCCTCTAACTCAAG
CTGACGTGACTTCAGCTAACCTATTCCAATCTATGTCACTAATTGTTGG
TGGTGCAGCTGTGGTATTAAATTGCCTTGATGAAATTAGTTAATGATACT
CCTGGAACAGCGACACCAATTGCAGGATTGCTGTATGTTGCTATAAC
CCAATGATAAAAGTACTAATAACCGCTCTAGGTGTATTATCCTATCTTA
CTAGCAGGCTATTGGAGGCATTGTTT

MWPNCAPLINSTLFTIEDILTSGAHSNPILMGVILGGTIVVVATAPLSSMALTA
MLGLTGMPMAIGALSVFGSSFMNGVLFHKLKLGSRKDNIAFAVEPLTQADVT
SANPIPIYVTNFVGGAACGILIALMKLVDTPGTATPIAGFAVMFAYNPMIKVL
ITALGCHIISLLAGYFGGIVF

Sequence description:

- A] Length: 540 bp - 180 aa (partial sequence)
- B] N- and C-termini have yet to be elucidated

ID-143**FIG. 1** CONT'D

Clone 2-c80

ATGTTTTAAGTATAATGGCAGGTGTCATAGCATTGCTCTGACAGTTATT
 GCCATTCCACGCTTCATTAAGTTTACCAATTGAAGAAAATTGGCGGGCAA
 CAAATGCATGAAGATGTCAAACACATCTAGCCAAAGCAGGTACGCCGAC
 AATGGGAGGAACGGTATTTT

MFLSIMAGVIAFVLTVIAIPRFIKFYQLKKIGGQQMHEDVKQHLAKAGTPTMG
 GTVF

Sequence description:

- A] Length: 172 bp - 57 aa (partial sequence)
- B] Shine Dalgarno sequence precedes 'ATG' start codon. Possesses a potential leader peptide sequence.

ID-144

Clone 3-83

ATGAAACCATAATTATCTTTATTGGTAGAACGTTATTATACTTCGGTATTT
 TATTGTTACTAATTACTTTTGCAACCTTGGTCGGACAAGGCAGTTT
 TATTATAAA

MKPYLSFIGRTLLYFGILLLIYFFAYLGRGQGSFIY

Sequence description:

- A] Length: 113 bp - 37 aa (partial sequence)
- B] Putative ATG start codon is preceded by a typical Shine-Dalgarno sequence. Possesses a potential leader peptide sequence.
 This orf is not in frame with nuc

FIG. 1 CONT'D

ID-145

Clone 3-86

ATGTCATATTTAGAAATTACTGGTATCGTTGGAGCAATTATTTATT
 TTTAGCAGTAATATTGCTTGTAGACCTGACTGGTCAATGCTTCACTA
 TCTATTGTATTTACTTTATGGCACTCTAGCGCATCAATTGAAGAATAT
 CAGTTCCCGGTGGGCATCACCTATCATTAACATGTTGTTATGATGAA
 GAAGAGCTGATGGATTGTTCCAGGCAAACTCAGTCTATTATGTTGGTT
 AATACTATTGCTTGGTTGCTTACATTGCTAGTATTGCTTTCCTCAAGCTT
 ATTGGCTTGGATTAGGAGTCATGTTCTTAGTCTAACGCAGCTCTGGGTC
 ATGGTTTCAGATGAATATTAAACTTAAACTTGCTATAATCCTGGCTAG
 CAACGACAGTATTCTCTAGTACCAATAGCTTGCACATCTATCAAG
 CTAGTGCAGAAGGAATGCTCACTGGGAGATTGGCTAGGTGGTTATCA
 TGTTGATTGTCTGTACTAACTAGCATTATTGCACCTGTACAGCTATTGAA
 GGATAAGGAGACCAATTATATTAGTCCTGGCAAATGGACCCTTCA
 TAAGTCGTTAATTGTAAAGGATAAAAAAAATAA

MSYFRNYWYRGAILFIILAVILLVFRPDWSMLHYLLYFYFMALLAHQFEYQ
 FP GGASPIINYVVYDEEELMDCFPGNTQSIMLVNTIAWLLYIASIAFPQAYWLG
 LGVMFFSLTQLLGHGFQMNIKLKTWNPLGATTVFLLVPIACAYIYQASAEG
 MLTWGDWLGGFIMLIVCVLTSIIAPVQLLKDKETNYIISPWQMDRFHKVVNFV
 RIKK*

Sequence description:

- A] Length: 651 bp - 219 aa (full length gene)
- B] Putative ATG start codon is preceded by a typical Shine-Dalgarno sequence. Possesses a potential leader peptide sequence.

ID-146

Clone 3-c88

FIG. 1 CONT'D

ATGCCACTTACAGCACTTGAAATTAAAGATAAAACATTTCATCAAAATT
CGCGGTTATAGCGAAGAAGAGTT

MPLTALEIKDKTFSSKFRGYSEEEV

Sequence description:

- A] Length: 75 bp - 25 aa (partial sequence)
- B] Putative ATG start codon is preceded by a typical Shine-Dalgarno sequence. No leader peptide

ID-147

Clone 3-90

ATGTCACTTTCAAGAAAAATTGCTTACAATTGCGCTAAAAGGAAGCG
CTTATAAAGAGAGTTAGGACGCTACGCCTTGAGATCAATGCTAGCAGG
GGCTTATTGACAATGAGTACTGCTGCCGTATCGTCGAGCTGATACTAT
TGGTAAAATTCTCCTGCTCTATCAGGTTTGTATTGCTTTCATCTTAGTT
TTGGACTTATTATGTTTAATATTAAATGGTGAATTGGCGACATCTAATAT
GCTTATCTCACTGCAGGAGCCTATAATAAAATATCTCTGGAAAAAGC
CATAACAAATTAAATTATTGTACTTTCAACCTCGTGGTGCTTGTATA
TTAGCTTGGTTGTTAA

MSLFQEKIAYNCAKKEALYKESLGRYALRSMLAGAYLTMSTAAGIVAADTIG
KISPALSGFVFAFIFSFGLIYVLIFNGELATSNMLYLTAGAYKNISWKKAITILI
YCTFFNLVGACILAWLF

Sequence description

- A] Length: 406 bp - 125 aa (partial sequence)
- B] Putative ATG start codon is preceded by a typical Shine-Dalgarno sequence. Possible leader peptide

FIG. 1 CONT'D

ID-148

Clone 3-92

AAGTTACAAGCGACTGAAGTTAAGAGCGTCCGGTAGCACAACCAGCTTC
 AACAAACAAATGCAGTAGCTGCACATCCTGAAAATGCAGGGCTCCAACCTC
 ATGTTGCAGCTTATAAAGAAAAAGTAGCGTCAACTTATGGAGTTAATGAA
 TTCAGTACATACCGTGCAGGGAGATCCAGGTGATCATGGTAAAGGTTAGC
 AGTTGACTTTATTGTAGGTAAAAACCAAGCACTGGTAATGAAGTTGCACA
 GTACTCTACACAAAATATGGCAGCAAATAACATTCATATGTTATCTGGCA
 ACAAAAGTTTATTCAAATACAAATAGTATTATGGACCTGCTAATACTTG
 GAATGCAATGCCAGATCGTGGTGGCGTTACTGCCAACCACTATGACCACGT
 TCACGTATCATTTAA

KLQATEVKSPVVAQPASTTNAVAHPENALQPHVAAYKEKVASTYGVNEF
 STYRAGDPGDHGKGLAVDFIVGKNQALGNEVAQYSTQNMAANNISYVIWQQ
 KFYNSNTNSIYGPANTWNAMPDRGGVTANHYDHVHVSF

Sequence description

- A] Length: 419 bp - 139 aa (partial sequence)
- B] N- and C-termini have yet to be determined

ID-149

Clone 3-94

ATGATTCCAGTAGTTATTGAACAAACAAGTCGTGGTAGAACGTTCTTATGAT
 ATTACTCACGTCTTAAAAGATCGTATTATTATGTTGACAGGCCAAGTT
 GAGGATAATATGCCAATAGTATCATTGCACAGTTATTGTTCTCGATGCA
 CAAGATAATACAAAGGATATTACCTTATGTCAATACACCAGGTGGTTCA
 GTATCGGCTGGACTTGCTATTGTGGACACCATGAACCTCATTAAATCGGAC
 GTACAGACGATTGTTATGGGGATGGCTGCTTCGATGGGAACCATTATTGCT
 TCAAGTGGTGTAAAGGAAAACGTTTATGTTACCGAATGCAGAATATG

FIG. 1 CONT'D

ATCCACCAACCAATGGCGGAACAGCGGAGGTACACAGCAATCTGATAT
 GGCTATCGCTGCTGAGCATCTTAAAAACGCGTCATACTTAGAAAAAAT
 CTTAGCTGATAATTCTGGTCAATCTATTGAAAAAGTCCATGATGATGCAGA
 GCGTGATCGTTGGATGAGTGCTCAAGAACACTGATTATGGCTTATTGAT
 GCTATTATGGAAAATAATAATTACAATAATAGATTAAAAGAGTTGAGTT
 TACCAACTCTTTTATTGTTGAATTATGTTATAATCTTAGTAATTACA
 GATATGACGCAGAAAGGAAAAATTATTGA

MIPVVIEQTSRGERSYDIYSRLLKDRIIMLTGQVEDNMANSIIAQLLFLDAQDN
 TKDIYLYVNTPGGSVSAGLAIVDTMFNIKSDVQTIVMGMAASMGTIASSGAK
 GKRMLPNAEYMIHQPMGGTGGGTQQSDMAIAAEHLLKTRHTLEKILADNSG
 QSIEKVHDDAERDRWMSAQEHЛИALLMLWKIIIYNNRFKRVEFTNSFFICW
 NYVILVITDMTQKGKNY*

Sequence description

- A] Length: 693 bp - 231 aa (full length gene)
- B] Putative ATG start codon is preceded by a typical Shine-Dalgarno sequence. No leader peptide. Significantly, it would appear to have a very hydrophobic C-terminus.

ID-150

Clone 2-c86

ATGAAACCAAAATATTGGTGTACTGGCTAGGAATATTGGACAAACA
 CTCGCACAAGAACTAAGTAACCTTGAACAAAGATGTTATTGCTATTGACAGC
 AATCCTGAAAATGTACAAGCTGTCGCCGAAGT
 TGTACAAAAGCAGCTATCGGAGACATTACTGATTAGCTTCTAAACAC
 CATCGGGATCAGTGACTGTGATACTGTTATTATTGCTACAGGAAACAGTT
 AGAGAGCTCAGTATTGCCGTAAATGCACTGTAAAAGTTAGGCGTCCCAC
 AAGTTATTGCTAAAGCTGAAACCTGTATACGAAGAAGTACTTTATGAAA
 TTGGTGTGATTGGTTATCTCTCCGGAGCGAGAATCTGGCAAAATGTTG
 CTGCAAACCTCATGAGAAATAAAATTACAGATGTCTCCAGATTGAATCTG
 ATATTCTGTCATTGAATT

MKPKIIGVLGLGIFGQTLAQELSNEQDVIAIDSNPENVQAVAEVVTKAIGDI
 TDLAFLKHIGISDCDTVIATGNSLE

FIG. 1 CONT'D

60 / 110

SSVLAVMHCKLGVPQVIKARNLVYEEVLYEIGADLVISPERESGQNVAAN
LMRNKITDVFQIESDISVIEF

Sequence description:

- A] Length: 459 bp - 153 aa (partial sequence)
- B] Putative ATG start codon is preceded by a typical Shine-Dalgarno sequence. Possesses a potential leader peptide sequence.
This orf is not in frame with nuc

ID-151

Clone 2-c88

GTGCGTTAGTAAAGAGATTATTCAGTTAGCTATACCAAGCTATGATTGAA
AATATCTTACAAATGCTCATGGGAGTAGTTGATAATTATCTAGTGGCTCAG
TTAGGTGTTGAGCAGTATCAGGTGTTCAAGTTGCTAATAATATAATTACT
ATTATCAAGCTATTTTATAGCTTACAGGGCGAGTATAGCAAGTCTATTG
GCCAAGTCGTTAGCAGGTAGTGAGAAGGGATGCAATTCAAGTATGTTCT
CAAGCCATTCTAACATCACTGATAGGGCAGTATTAGGAATTATCTCG
ATTGTTTTGGACAAACTTCTT

MRYSKEIIQLAIPAMIENILQMLGVVDNYLVAQLGVVAVSGVSVANNIITY
QAIFIALGASIASLLAKSLAGSEKDDAISVCSQAIFLTSLIGAVLGIISIVFGQTFF

Sequence description

- A] Length: 330 bp - 110 aa (partial sequence)
- B] Putative GTG start codon is preceded by a typical Shine-Dalgarno sequence. May have a leader peptide

ID-152

FIG. 1 CONT'D

Clone 2-c92

TTGATTAACAAGTATCGTCTTTGAAGAGGATTCTCCATAATAACT
 CCTTAATAGTTATCGTGAGAAGTATTAAAGAAAAACGCCAAGGTAG
 AGCGACATTCTGCCTTAACATAAAAACCAAGAGAATTAGCACAAC
 ATTATCTCTAAAATTACAAAGTTCTCAAGGGTTTAGGAATAGCTAGTG
 AATTGGTAACCTATGATCAACGCTTGTCAAACATTTT

MINKYSCFLKRILHNNTPLIVIVRSILKKNRQGRATFLPLTTIKPRELAQHYLSK
 LQSSQGFLGIASELVTYDQRQLSNIF

Sequence description

A] Length: 240 bp - 80 aa (partial sequence)

B] No obvious Shine Dalgarno sequence precedes the Putative TTG start codon

ID-153

Clone 2-c94

TTGTTGACTCACAAAATATTAAACCATTATTTGGATTATTTATGA
 TTATATTATCAGCATGTGGTATGTCTAATAAGGAAATGGCTGGTATTGATA
 ATTGGGAACATTATCAAAGGAAAAGAAAATTACTATTGGATTGATAAT
 ACTTTTGTCTATGGGATTGAAAGTCGTCTGGTACTATACCGGCTTG
 ATATTGATTAGCTAATGCTTTAAAGAACGGTATTCACTGAAAT
 GGCAGCCTATTAACGGGATATGAAAGAAACTGAACCTAATAATGGTAAT
 ATAGACCTTATTGGAATGGTTATTCAAAAACGGCAGAACGTGCTAAAAA
 AGTCGCTTTACAAACCCATATGAATAATCATCAAGTAATTGTTACTAA
 AACTTCATCACATATTAAATAGTATTAAAGGATATGAAGGGGGAAAAACTAG
 GAGCCCAGTCGGGTTCATCTGGTTTGATGCTTTAACGCTAACCTGATA
 TTTAAAAAAAGTTGAAAAGGAAAAGAAGCAGTTCAATACGATACTTTC
 ACTCAGGCTTGATTGATTAAAAAATAACCGTATTGATGGTCTTGATT
 GATGAAGTTATGCTAACTATTAAAGCAAGAAGGAA

FIG. 1 CONT'D

MLTHKNILLTIIFGLFMIILSACGMSNKEMAGIDNWEHYQKEKKITIGFDNTFV
PMGFESRSGDYTGFDIDLNAVKEYGISVKWQPINWDMKETELNNGNIDL
WNGYSKTAERAKVAFTNPYMNHHQVIVTKTSSHINSIKDMKGKLGQAQSG
SSGFDAFNAPDILKKFVKGKEAVQYDTFTQALIDLKNNRIDGLLIDEVYANY
YLKQEG

Sequence description

- A] Length: 649 bp - 216 aa (partial sequence)
- B] TTG start codon is preceded by a possible typical Shine-Dalgarno sequence. Has a leader peptide

ID-154

Clone 2-c100

ATGAAAATTGGAAAAAAATAACCTTAATGTTTCTGCAATTATTTAACA
ACAGTAATTGCATTGGGAGTCTATGTTGCCTCAGCTTATAATTTTCGACTA
ATGAATTGTCTAACAGACTTT

MKIWKKITLMFSAILTTVIALGVYVASAYNFSTNELSKTF

Sequence description

- A] Length: 123 bp - 41 aa (partial sequence)
- B] ATG start codon is preceded by a potential typical Shine-Dalgarno sequence. Has a typical leader peptide

ID-155

Clone 2-c1

FIG. 1_{CONT'D}

ATGAAAAAAACAAAGACTATTACTGCTTTGGAGGCTTATTAATAATGATA
 ATGATGACAGCATGTAAGGATTCAAAAATCCCAGAAAACCGCACGAAAAAA
 GGAATACCAGGCAGAACAGAATTAAAGTCATACTTTAAATATATATCAG
 ATAAAAAATAACTATTAGATAATATAAAAGTTATTACTTTCTATAAGTA
 TTCTAAAGATGTACAAGATAAAGTCAGTGAAACAACAACTGTTCATATA
 GACTAGAAAAGCAAAAGAATCAAGAGTTCAAGAGTTCAAGAGTTAATTGAACATGAA
 GTTAGTGAATCTAGTCATATTCAACCGAAGTTAAAATCAAATACAGTAT
 CCAATCCAGTATAAAGATAATTCAATTGTTACTGAAAAAACACCGTCA
 GAACGTTATGATGAGTTGTTAGTTCAATTGATTCTTCAATTAAATTTA
 AATATAAAATATGATTACTAAACATCCCAGAACTGAATTAAAAA
 GGTGTTCTATAAGATTCTATAAATTCTGAAATTGTAGCCCCTTTATAA
 ATCAATTAAATATAAAATCCTAAAAATCATCTATTGGTTACAAAAAA
 CGGAAAGTAAAGAATATTATTACAAATCAGTATTGATACTGATTCTGAGA
 TATATTCTATATTCGAAGGTATTCA

MKKQRLLLFGGLIMIMMTACKDSKIPENRTKEYQAEQNFKSYFKYISDKN
 NYLDNIKVYYFSISISKDVQDKVSETTCSYRLEKQKNQEFIGNEHEVSESSQ
 YSTEVKNQIYPIQYKDNSIRFTEKTPSERYDEFVFSSFDSSLKKYKIYDYLK
 HPETELKGVSYKIPINSEIVAPPINQLNIKNPKKSSISVTKTESKEYYYTISIDTS
 EIYSIFEGIH

Sequence description

- A] Length: 687 bp - 229 aa (partial sequence)
- B] ATG start codon is preceded by a potential typical Shine-Dalgarno sequence. Has a typical leader peptide. C-terminus has yet to be verified

ID-156

Clone 2-c5

ATGACATTGACACCATTGATCAATTAGCGGTTAACAGTCGCACGCTT
 TCTATTGATGCTATCCAAGCAGCAAATTCTGGGCACCCAGGTCTTCTATG
 GGAGCTGCGCCTATGGCTATGTGCTTGGAAATAATTCTAAATGTAAC
 CCAAAAACAAGTCGCAATTGGACAAACCGTGACCGTTTGTACTTCAGCT

FIG. 1 CONT'D

64 / 110

GGGCATGGTTCAGCTCTTCTTATGCCACTTCATTTAGCTGGCTATGATT
TATCAATTGATGATT

MTFDTIDQLAVNTVRTLSIDAIQAANSGHPGPLPMGAAPMAYVLWNKFLNVNP
KTSRNWTNRDRFVLSAGHGSALLYSLHLAGYDLSIDD

Sequence description

- A] Length: 272 bp - 90 aa (partial sequence)
- B] ATG start codon is preceded by a potential typical Shine-Dalgarno sequence. No obvious leader peptide

ID-157

Clone 2-c8

ATGAGAACACTATTTAGAATGATATTGCTATTCAAAGTTATCTTTAGA
TTGATTGGAATATCATTTGGGAATTCAAGACAGTTCTTGTATTGCG
ATTATTTATTTGGCTTGTATTACTATGCGAACATCACAGTCAATCAGAATTG
CTAATCAACTTAGTGACATTATTCAGACAGGAAAAACATTTT

MRTLFRMIFAIPKFIFRLIWNIIWGIFKTVLVIAIILFGLYYYANHSQSEFANQLS
DIIQTGKTF

Sequence description

- A] Length: 197 bp - 65 aa (partial sequence)
- B] ATG start codon is preceded by a potential typical Shine-Dalgarno sequence. Possesses a leader peptide

ID-158

FIG. 1 CONT'D

Clone 2-c9

ATGTCAAAAAAAATAATATTAGGAATTTCATCTCTTATCTGTCGTTACTT
TGGTGGCGTGTGGTCATCAGACAAACAGCTACAAGATAAAGTTGAGAAA
AAAGGGAAGTTAGTTAGCGGTGAGTCCAGATTATGCTCCCTTGAGTT
MSKKIILGILSLLSVVTLVACGSSDKQLQDKVEKKGKLVLAVSPDYAPFEF

Sequence description

- A] Length: 153 bp - 51 aa (partial sequence)
- B] ATG start codon is preceded by a potential typical Shine-Dalgarno sequence. Possesses a leader peptide (not in frame with nuc)

ID-159

Clone 2-c10

ATGAAAATCAAAGACTATTACTGCTTTGGAGGCTTATTAATAATGATA
ATGATGACAGCATGTAAGGATTCAAAATCCCAGAAAACCGCACGAAAAAA
GGAATACCAGGCAGAACAGAATTAAAGTCATACTTT
MKNQRLLLLFGGLLIMIMMTACKDSKIPENRTKKEYQAEQNFKSYF

Sequence description

- A] Length: 139 bp - 46 aa (partial sequence)
- B] ATG start codon is preceded by a potential typical Shine-Dalgarno sequence. Possesses a leader peptide

FIG. 1 CONT'D

ID-160

Clone 2-c11

ATGATTGGAAAATTATATTAGCTATAGAAAGTCACGCTTATTAAGAAGT
 ATTTATGGCTTATTTAATTGTTGGTGTATATGTTAGGACAACGTGTTT
 TATTATCCACTGTTCTTATCACATCAAGAGATAAAACTAGCAGTAGATC
 AACATTTACTCAATAACTTTCAAGCAGTAAGTGGTGGAGTTAATAAAAT
 TAAATGTTTACACTGGGTTGACTCCATGGATGTCAAGTATGATTATTT
 GGAGATTCGTTCTTATTTCGTGGCAAAAAATGCAACGAAGCGAAAAA
 GCAGAAGTAGCTCAATATACTTTAATGCTTACTATCTCAGTTACAGCA
 TATGGTGTTCAGGAAATCAATTATAAAAAGCTCTTATTAGGTTCTTATA
 GTGATATTGTTTT

MIGLYYSYRKSRLLRSILWLILVGVYMLGQRVLLSTVPLSHQEIKLAVDQHL
 LNNFSAVSGGSFNKLNVFTLGLSPWMSSMIIWRFVSLFWAKNATKRKAEVA
 QYTLMLTISVIQAYGVSGNQFIKSSLGGSYSDIVF

Sequence description

- A] Length: 423 bp - 141 aa (partial sequence)
- B] ATG start codon is preceded by a potential typical Shine-Dalgarno sequence. Possesses a leader peptide

ID-161

Clone 2-c13

ATGAAAGGTCTATTGGATTTTAGTTAATATTGCCAGAACGCCAGCTATT
 TTAGTCGCCTTGATAGCCATTATCGGTTAGTACTGCAGAAAAAAGGTGTT
 CCTGATATTGAAAAGGTGGAATAAAAACATTGTTGGCTTCTTAGTGGTT
 TCTGAAGGTGCAGGGATAGTCCAAAATTCTGAATCCATTGAAAAATG
 TTTGAACATGCTTTCATTGGTGGGGTAGTTCTTAATAATGAAGCCATT
 GTAGCAGTAGCTCTACGAAGTATGGCTCAGCAACTGCTTGTATTGTTA
 GCAGGAATGATTTAATATTGCTGTTACAAAAA

FIG. 1 CONT'D

MKGLLDFLVNIARTPAILVALIAIIGLVLQKKGVPDFVKGGIKTFVGFLVVSEG
AGIVQNSLNPFKGKMFHAFHLGVVPNEAIVAVALTKYGSATALIMLAGMI
FNILIARFTK

Sequence description

- A] Length: 348 bp - 116 aa (partial sequence)
- B] ATG start codon is preceded by a potential Shine-Dalgarno sequence. Possible leader peptide

ID-162

Clone 2-c21

TTGGTTGGTAAGCCCCAATTACTATTTTAGATGAACCTACTTCCGGAATG
GATACTTCCACACGTCAACGATTTGGAAGCTGGTTGCACACTAAAAAA
AGAACGGTGACACAATTGTCTATTCTAGTCATTATATCGAAGAGGTAGAAC
ATACAGCTGATAGGATTTAGTACTTCATAAAGGAAAGTTATTACCGCATA
CAACCCCCCTTGCATGAAGCAAGAAAAACCGAAAAGTTATTCCACCGTT
CCGCTTAGTTATCAAAAATTATTACCTACCTATTGATTACAGAGTGTGAA
GCCAAGAGTGTAGTATAACGTTGTTACTGGGGAGGCTGAAACTGTATG
AAAAAACTGGCAGATAATGGTTGTCTATTGAAGCTATTGAGATGACCA
ATAGAACTTGTAAATCGTATTTTGAGACTACTAAGGAGGTAAAACATG
AGAATCTTA

MVGKPQLLFLDEPTSGMDTSTRQRFWKLVATLKKEGDTIVYSSHYIEVEHTA
DRILVLHKKGKLLRDTTPFAMKQEKTKEKLFTVPLSYQKLLPTYLITECEAKSDSI
TFVTGEAETVWKILADNGCPIEAIEMTNRTLLNRIFETTKEVKHENL

Sequence description

- A] Length: 462 bp - 155 aa (partial sequence)
- B] Putative TTG start codon is not preceded by an obvious Shine-Dalgarno sequence. No obvious leader peptide. N- and C- termini require further

FIG. 1 CONTD

examination.

ID-163

Clone 2-c25

TTGAAAAAAATCCAAGAGAAGCCGTAAGGCAGTGACAACAAGTGGTGAGA
AGACTTACTGAGGATTGGCAAAAATGAATTCTAGACGAAGTCATTA
ATGTTATGGTTTATATACCTTGAATAAGACAAAATCTGCTAACCTAAATA
AGGCCTATATCATGAAAGTTGCTAATGATTTCGCCTTCAGAATGTTATGA
CGGCCGAAGATGCTGTGCTTAAATTCGTGATTTCAGATCAAAAAGTAA
GGACTAAAACAGAACGAAGAAGAAACAATCGAATGTTCCCTGAATGGAGT
AATCCTGATTATAAAGATGAGGTTAGCCCAGAAAAAGAAATTGAATTAGA
ACAGTTT

MKKSKRSRKA VTTSGEKTLEDLAKMNFLDEVINVMVLYTLNKTSA NLNK
AYIMKVANDFAFQNVMTAEDA VLKIRD FSDQKV RKTETKKQSNVPEWSN
PDYKDEVSPEKEIELEQF

Sequence description

- A] Length:360 bp - 120 aa (partial sequence)
- B] N- and C- termini require verification.

ID-164

Clone 2-c28

ATGACGAATCATATTACTAAACTGATAGAAAATAGCGGAAAAAAATTGAC
AGAAATTAGCGAAGCTACAGATATAGCCTATCCTACACTTCTGGATACAA
TCAAGGAATCCGAAACCTAAAAAAAGATAATGCTGAAAAATTGGCAAAAT
ACTTAATGTTCCGTCGCTTACATTATGGGACTTGATAGCAACCCACATG
CTCCATCAAATCTT

MTNHITKLIENSGKKLTEISEATDIA YPTLSGYNQGIRKP KKDNAEKLAKYFNV
SVAYIMGLDSNPHAPSNL

FIG. 1_{CONT'D}

Sequence description

A] Length:218 bp - 72 aa (partial sequence)
B] ATG start codon is preceded by an
obvious Shine Dalgarno sequence. No obvious
leader peptide.

ID-165**Clone 2-c29**

TTGATGAAAAGGAATAAACATTACCGTTAACAGAAACTACCTATTATATT
TTATTAGCTTGTGAGGAAGCGCATGGCTATGCTATTATGAAAAAAGTT
GAAGAAATGAGTGGCGGTGATGTTAGAATAGCCGCAGGGACAATGTACGG
TGCCATTGAAAATTTACTTAAACAAAAATGGATAAAGTCTATCTCAAGTGA
CGATAGAAGAAGAAAAGTTATATTACTGAGACAGGAAAAGAAAATAG
TAGAACTTGAAACGAATCGATTAAGAAAGTTACTTAATACTGCTAATCAGT
TGGGTTTGGAGGAGATGGTTATGATAAAGTT
MMKRNKHLPLTETYYILLALFEEAHGYAIMKKVEEMSGGDVRIAAGTMYG
AIENLLKQKWIKSISSDDRRRKVYIITETGKEIVELETNRLRKLNTANQLGFG
GDGYDKV

Sequence description

A] Length:337 bp - 112 aa (partial sequence)
B] TTG start codon is preceded by an
obvious Shine Dalgarno sequence. Actual start
codon may ATG that comes immediately after the
TTG. Potential leader peptide.

ID-166**FIG. 1 CONT'D**

Clone 2-c35

CCCATTAAGTGGTGAAGTTAATAGCTGAGAAATTAGGAGTACCAAGAGCAGC
 ACTAAGGTCTGATTGCGGGTTTAAGTATGCTAGGTATCATAGATGAAA
 ACCTAAGGTGGTTATTTTATTTAGGACAGTATCATGCTCAATAGGGAC
 AAGTCATTGAAAAGATGACAGTTCAGAAATTATGGGGATCCCTCTGAC
 AGTTCATAAAAAGATTCAAGTTATGATGTTATTGTACATATTTATGGA
 AGATGCTGGTTGTGCTTTATCTTGGATGATGATGATTCTCTGTGGAGTC
 GTGTCACGTAAAGATTACTAAAACCAGTATTGGCGGAGGAGATCTTCT
 AAAATGCCAATAGGAATGGTATGACACGTATGCCACACGTGACAACGT
 TTAGAAAAATGAAAGTCTTTGCGGCAGCTGATAAAATTAGTGAGCAGAA
 AAGTGGATAGTCTCCCTGTCGTCATGATAAGCAATATCCGAAAAATT
 TTA

PITGELIAEKLGVPRALRSDLRVLSMLGIIDAKPKVGYFYLQYHASIGTS
 HF EKMTVSEIMGILLTVHQKDSYDVIVHIFMEDAGCAFILDDDFLCGVVS
 RKD LLKTSIGGGDLSKMPIGMVMTRMPHVTVLENESLFAAADKLVSRKVDSL
 PV VRHDKQYPEKF

Sequence description

- A] Length:511 bp - 170 aa (partial sequence)
- B] N- and C-termini to be determined

ID-167

Clone 2-44

TTGGAAGTCATCATGCAATTATTTAGTATTATTGGTATTTATTGGTAT
 TAGGAATTGTGTATGCAATTCTTCAATCGTAAGAGTGTCTCTAAGTTT
 ATTGGAAAAGCTTATCGTCAATTCAATTGCGCTAATCTTAGTACGT
 ATCCCACCTAGGCCAACAAAGTTGTTAGTGTGTTCAACTGGAGTTACTAAA
 GTAATCAACTGTGGTCAAGCTGGTT

MEVIMQFIYSIIGILLVLGIVYAIISFNRKSVSLSLIGKALIVQFII
 ALILVRIPLGQQ VVSVVSTGVTKVINCQAG

FIG. 1 _{CONT'D}

Sequence description

- A] Length:233 bp - 77 aa (partial sequence)
 B] TTG start codon is preceded by a possible Shine Dalgarno sequence. Actual start codon may occur further downstream. Potential leader peptide.

ID-168

Clone 2-46

CAACCTAATAAAAGCTTAGAAAGTGTGAGATTGATATTAAATGCTTCCAG
 CATTATAATTACTAACCAATTGGAATAAAGCAAATAAGACCAATCTGTT
 TCCGTTGCTGAGACATACTTTACTTCCTTAGATTATACTCTGGTACTAAGA
 ACGGTAAAGGTAACCAACAGTTCTGAAATTCCAATAAAGCAACT
 ATTACTATCCCACGATGCAGTTAACGAAAGTCGCTCTACTTGTAA
 CAATCAGCAGGCTGCTAAAATTGAAAGTATCAGGTGATACATTAGAAC
 AATGTCAGATGTTGTTCCAATCCTAAATCTTAGATT
 QPNKALESDEIDINAFQHNYLTNWNKANKTNLVSAETYFTSFRLYSGTKN
 GKGKYQTVSEIPNKATITIPNDAVNESRSLYLLQSAGLLKLKVSGDTLATMSD
 VVSNPKSLD

Sequence description

- A] Length:344 bp - 114 aa (partial sequence)
 B] N- and C- termini require verification

ID-169

Clone 2-47

ATGAAATGTATAATAATAATATAAAATAAAAAATGATAATTGAGAT
 TTATCATAGAAGGAAAATTTGAAATTAAATAAAATCATATTATCTAC

FIG. 1_{CONT'D}

72 / 110

TGCAGCTCTTACTGCTCTTTAGGATATAATAGCGTTACTGCGGATACA
 TATAATAACTATCAGCCACATAGATCAAATAATGGATTTAAGTGA
 ATATAACTATAATAACCAGATAGAACCTTCAGGAGCGTATAAAAAACCTAA
 ATATACCTTT

MKCIINNINKIKMIEIYHRRKTILKLNKILSTAALTALFLGYNSVTADTYNNY
 QPHRSNNMDLTEEYNYNNQIELQERIKNLNIPF

Sequence description

- A] Length:264 bp - 88 aa (partial sequence)
- B] There is a Shine-Dalgarno sequence upstream
of this sequence. Potential leader peptide
sequence

ID-169

Clone 2-47

ATGAAATGTATAATAATAATATAAATAAATATAAATGATAATTGAGAT
 TTATCATAGAAGGAAAATCTTTGAAATTAAATAAATCATATTATCTAC
 TGCAGCTCTTACTGCTCTTTAGGATATAATAGCGTTACTGCGGATACA
 TATAATAACTATCAGCCACATAGATCAAATAATGGATTTAAGTGA
 ATATAACTATAATAACCAGATAGAACCTTCAGGAGCGTATAAAAAACCTAA
 ATATACCTTT

MKCIINNINKIKMIEIYHRRKTILKLNKILSTAALTALFLGYNSVTADTYNNY
 QPHRSNNMDLTEEYNYNNQIELQERIKNLNIPF

Sequence description

- A] Length:264 bp - 88 aa (partial sequence)
- B] There is a Shine-Dalgarno sequence upstream
of this sequence. Potential leader peptide
sequence

FIG. 1 CONT'D

ID-170

Clone RS-58b

TTGGGTGATTATTATGGTAAGAAATATTTGGTGAGGCAGCTAAAAAAGA
 CGTCGAACATATGGCTAACGAAAATCATTAATGTCTATAAAACACGGTTAA
 AAAACAACACTGGTTATC
 AGAAAATACAAAAGCAATGCCATTAAGAAACTGATAACATGAGATTA
 TGATTGGCTATCCAGAAGATTATCCTGATCTTATCGTCAGTACCAATTG
 ATAGTAAAGCAAGCTCTTGAAAACAATGATAACTACAGAAAATTATCG
 AACAAAGAAAACATTGAAGAATTAAACCAGTCTAACACGTGAACATTG
 GCAAATGAGTGCCTATGCTGAAATGCTTATAATGATCCTAACACCAATT
 CATAGTCTTCCAGCAGCGATTTCATCACCAGTGTACGATAAAACTAA
 AACAGITAGTCAAAATTATGGAGCTATGGAGCAATTATTGGTCATGAAAT
 TTCACACTCATTTGATATTAAATGGTATGAAATATGACGAGAAAGGGAATCT
 TCACGATTGGTGGACTAAAGAAGATTAAATCATTATAAGAAATCAACAC
 AAGCTATGATTGACCAATGGGATGGCCTAAAGCAGATGGCGTAAAGTT
 GATGGTAAATTAACTTAGCAGAAAATTGCAGATAATGGTGGTGTATG
 GCATCTCTAGAAGCTTAAAGACTGAAAAAAATCCAAACTATAAAGAATT
 TTGAATCATGGGCAAGTATTGGCGTAAAAAGCAACCAAGAACAAAGT
 AAGCTCTCAATTCACTAGTCAGATGTTATGCACCATATGAATTGA>
 GAGCTAACATCCCAGTACGTAATTCCAAGAATTATGATGCCATTGGT
 TTAAAAAAGGCGATTCAATGTATCTAAAACCAGAAAAACGTTGACACTT
 GGTA

MGDYYGKKYFGEAKKDVEHMAKKIINVYKTRLKNNTWLSENTKAMAIKK
 LDNMRLMIGYPDYPDLYRQYQFDKASFFENNDNYRKLSNKKTFEEFNQSNQ
 REHWQMSANAVNAYNDPNTNSIVFPAAIFQSPLYDKTKTVSQNYGAIGAIIGH
 EISHSFDINGMKYDEKGNLHDWWTKEFLNHYKKSTQAMIDQWDGLKADGG
 KVDGKLTLAENIADNGGVMASLEALKTEKIQTICKNFLNHGQVFGVKKQPKNK
 VSPQFSQMFMHHMN*

Sequence description:

A] Length: 819 bp - 272 aa (full length gene)
 (107 bp of additional DNA sequence (> onwards) is
 also included. While not in-frame with the
 described orf, it also shares strong homology
 with the neutral peptidases.

FIG. 1 CONT'D

B] This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-89 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-89 gene sequence. ID-89 and ID-170 together show homology over their combined entire length with the neutral endopeptidases from *Lactococcus* and *Lactobacillus*. Possesses TTG (possible ATG start codon located 13 bp further downstream) start codon with no obvious signal peptide. Shine Dalgarno sequence not immediately obvious. Possibly located further downstream

ID-171

Clone 2-18/22b (Mod2)

```

ATGACCATGATTACGCCAAGCTTCATTAAGGTATCTCTAGATGAAACAAAT
CGTATGATGCGTATGATATCAGATTATAAGTTATCGCGCATTGATAAT
GAAGTAACGCATTTAGATGTTGAAATGACGAATTTACAGCTTCATGACC
TCAATTGAAATCGATTGATCAGATTAGAAATCAAAAAACAGTCACAGG
AAAAGTTATGAAATTGTCAGAGATTATCCTCTAAGTCAATTGGGTGGA
AATTGATAACAGATAAGATGACTCAAGTGATTGATAACATTAAATAATGC
AGTCAAGTATTACCCAGATGGTGGTAAGATTACAGTTAATCTACGCACAAC
TAAAACGCAGATGATTITATCAATATCAGACCAAGGCTTAGGTATTCCCAA
AAAAGATTACCTCTCATTTGATCGTTTTATCGTGTGATAAGGCGAGA
AGTCGTCAACAGGGTGGGACTGGACTTGGTTGTCAATTGCAAAAGAAAT
TGTTAAGCAGCATAAGGGATTATTTGGGCTAAGAGTGAAGTGTAAAG
GGTCTACTTTACAATCGTCTTGCCTATGATAAAAGATGCTGTAACITATGA
AGAATGGGAGGACGTTGAAGATTAA

```

```

MTMITPSFIKVSLDETNRMMRMISDLLSLSRIDNEVTHLDVEMTNFTAFMTSIL
NRFDQIRNQKTVTGKVYEIVRDYPLKSIWVEIDTDKMTQVIDNILNNAVKYSP
DGGKITVNLRTTKTQMILSISDQGLGIPKKDLPLIFDRFYRVDKARSRQQGGTG
LGLSIAKEIVKQHKGFIWAKSEYKGSTFTIVLPYDKDAVTYEEWEDVED*

```

Sequence description:

- A] Length: 613 bp - 212 aa (full-length gene possibly)
- B] Possible Shine Dalgarno sequence present upstream of a ATG start codon. May not have yet determined the N- portion of this gene. No obvious signal peptide.

FIG. 1 CONT'D

SUBSTITUTE SHEET (RULE 26)

ID-172

Clone 2-54balternate (107b)

TTGAAAAAAATTATTACTTCTATTCTATTACTTAGTTGCATTTTTTATGC
 CAACCATCTGCTGAATCTTTAATGCTTCCGCTAAACATGCCTTAGCAGT
 TGATTAGATTCAAGGAAAAATCTTGTATGAAAAAGATGCTAACAAACCG
 CTGCTATTGCTTCCTTGACTAAAATAATGACCGTTATATGGTCTATAAAG
 AAATTGATAACGGTAACCTCAAGTGGAAATACCAAAGTAAATATCTGAC
 TACCCCTTATCAACTAACACCGGAATCTGATGCTAGTAATGTTCTTAGAA
 AAAAGGCGCTATACTGTTAAACAACCTCGTGGACGCTGCCATGATTCTAGT
 GCTAACAGTGCAGCATTGCTTAGCTGAACATATTCAGGAAC TGAAAGT
 AAATTGTTGATAAAATGACTGCTCAATTGGAAAAGTGGGGAAATTGATGAT
 AGCCACCTAGTCATGCTTCTGGCTTAAATAATAGTATGTTAGGCAATCAC
 ATTTATCCAAAATCGTCACAAAACGACGAAAATAATGAGTCACGTGA
 TATTGCTATTGCTGCCTACCATTGGTCAACGAATATCCTCCATTCTTAAG
 ATTACTAGTAAGTCCGTTGCTAAATTGATAAAGATATTATGCATTCTTAT
 AACTACATGCTACCAAGATATGCTGCTTAGACCAGGTATTACAGGTTG
 AAAACTGGACAACGGAATTAGCTGCCAATCTTATTGCTACATCTACT
 GAAAGTGGAAATGAGACTACTCACTGTTATTATGCATGCTGATAAGGCCGAT
 AAAGACAAATATGCTGCTTACAGCAACTAACTCTCTGAACATATC
 ACAAAACACCTACGAACCTAACCTTGTATTAGCTAAAGGAGCTGCATATAA
 AGGTAAAGAAGCAAGTGTGAGAGACGGAAAAGAACAAATCGGTACATCGCT
 GTTGCCTAAACGATTGAAAGTAGTACAGAAGAAAATACTAAACA
 AAATCAGTTAAAATTAACCTTAAAGAGCTTACTGCTCTTATTACAAA
 AAAAGAGAACCTAGGGAAAGCTTATTACGTTGACCTTAATAAGGTTGGAA
 AAGGCTATCTCATAAAGAACCTAGCGTTCATTTAGTGGCAAAAGATAGT
 ATTGAGCGCAGTTCTTCCTCAAAGTGTGGTGAATCATTGTGCGCTAC
 GTTAACGAAAAACTTAA

MKKIITSILLSCIFFMPTISAESFNASAKHALAVDLDGKILYEKDANKPAAIA
 SLTKIMTVYMVYKEIDNGNLKWNTKVNI SDYPYQLTRESDASNVPLEKRRYT
 VKQLVDAAMISSANSAAIALAEHISGETSKFVDKMTAQLEKWIHDShLVNA
 SGLNNMSLGNHIYPKSSQNDENKMSARDIAIAAYHLVNEYPSILKITSKSVAKF
 DKDIMHSYNMLPDMPVFRPGITGLKTGTTELAGQSFIATSTESGMRLLTIVM
 HADKADKDYARFTATNSLLNYITNTYEPNLVLAKGAA YKGKEASVRDGKE
 QSVIAVAKNDLKVVQKKNITKQNQLKINFKELTAPITKENLGKAYVDLN
 KVGKGYLIKEPSVHLVAKDSIERSFFLKVWWNHFVRYVNEKL*

FIG. 1 CONT'D

Sequence description:

- A] Length: 1236 bp - 412 aa (full-length gene sequence possibly)
- B] A possible Shine-Dalgarno sequence precedes the putative 'TTG' start codon. (needs further cloning and sequencing to verify N-terminus)

ID-173

Clone 3-60b

ATGACGCITCGAGAATTACAATAGAAGAACATTAAAGAACATTCAAGGAAA
 TTATGATTACAATCATTTCACAAACACCTGAGATGGCTAAACTTTAGA
 AAAACGCGGCTATGATGTTAGGTATTGGGATATCAAGTAGAAAATAAAC
 TAGAGATAATCAGTTATCTTATATTATGCCAGTCAGTGGTGGTTCAAAT
 GAAAATTGATTCAAGGACCAGTCATTCAAATTCTAAGTATCTAAAACAATT
 TTATAAAGCATTGCAAGGCTATGCCAAATCCAACGGTGTCTAGAATTAAT
 AGTGAGCCTTTGATGATTACCAATTTCAGTTCAGGGAGTTCCCTAGT
 AACAGGGAAATGATAATCTGATTGAAGATTTACCAAGTTCAAGGTTATCAC
 CATGATGGTTAACAACTGGTTTACTGGTAAATATTATCTGGCACTATG
 TAAAAAATTAGAAGGTGTCACTTCTGAAACGTTACTATCTTCATTCTCTAA.
 GACAGGACGAGCTTGGTAAGAAAGCAATGCTTTGGAATCAAGGTC
 GCGTTCTAACGTGATGAGCTACATTATTAAAGAGATAACAACTTCTA
 CGTCAAATAGACGTGATTATGGATAAGTCCTTAGATTATTATCAAGATT
 TTTACGATAGCTTGAAGGCAAGGCTGAATTGTGATTGCCACTTAAATT
 TTAGAGAATACGACCATAACTGCAAATAAGCTGAAGCATTGGAAAAT
 AAGCTT

MTLRELTIIEFKEHSGNYDSQFLQTPEMAKLLEKRGYDVRYLGYQVENKLEI
 ISLSYIMPVTGGFQMKIDSGPVHSNSKYLQFYKALQGYAKSNGVLELIVEPF
 DDYQLFTSSGVPSNQGNDNLIEDFTSSGYHHDGLTTGFTKYLSWHYVKNLE
 GVTSETLLSSFSKTGRALVKKAMSGIKVRVLKRDELHLFKEITTSTSRRDY
 MDKSLDYYQDFYDSFEGKAEFVIATLNFR俞DHNLQIKAEALENKL

Sequence description

- A) Length: 771 bp - 257 aa (partial gene sequence)
- B) This gene sequence was not identified using the LEEP system. It was identified immediately downstream of the ID-65 gene which was identified by

FIG. 1 CONT'D

LEEP, during cloning and sequence analysis of the full-length ID-65 gene sequence. Sequence Characteristics:
 No obvious leader peptide sequence
 Orf is preceded by a potential Shine-Dalgarno sequence.

ID-174**Clone 2-17b (ID-80b)**

TTGTCATTAAGTTGGTGCAGTGTAAATCTTATCCCTCCTAAAATCATGG
 GATCAGTTATTGATGCTATTACAACGGAAAATTAAACAAGACCACAATTAC
 TATGGAATTATTAGGTTGGTTGTCAAGCTTAGCTATGTATGGGCTGCG
 TTATATTTGGCGTATGTATATTAGGGACTTCTTACAAATTAGGCCAAGTT
 GTCAAGATAACCGTTATTGAACATTACAAAAATGTCTCCTCTTTTATC
 AGAAATATCGTACAGGTGATTAAATGGCGCACCGCACCAACGACATCAAT
 TCTCTAACACCGCTTGCAGGAGGAGTTATGTCAGCAGTGGATGCCTCT
 ATCACAGCATTAGTAACGCTTATCACCATGTTTACTATTTCGCGCAA
 ATGACATTAATTGCGGTTATCCCTTGCCCTTAATGGCCTAGCACTAGTA
 AATTGGGGCGAAAAACCCATGAAACCTCAAAGAATCTCAGGCAGCCCTT
 TTCAGAATTAAATAATAAGTG

MSLSLVAVNLIPPKIMGSVIDAITGKLTRPQLLWNLLGLVLSALAMYGLRYI
 WRMYILGTSYKLGQVVRYRLFEHFTKMSPSFYQKYRTGDLMAHATNDINSLT
 RLAGGGVMSAVDASITALVTLITMFFTISWQMTLIAVPLPLMALALVNWGEK
 PMKPSKNLRQPFSELNNKV

Sequence description

A) Length: 534 bp - 178 aa (partial gene sequence)

B) This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-80 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-80 gene sequence.

Sequence Characteristics:

No obvious leader peptide sequence

Orf is preceded by a potential Shine-Dalgarno sequence.

FIG. 1_{CONT'D}

ID-175

Clone 2-11Ab (ID-103b)

ATGCATATTGAGACTGTTATTGATTCAAAGAATTAGGAAAAAGATATCGT
 TTTAAAATCCTACAAAAGAATTATAGCTGATACTTTAGAACAAAGTCTTA
 GAAGTGATAAAAGAACGGTATTATTATCAATCTCAAAATTATTATGTTGTT
 GGTTATTATCTTATGAAGCATCTGCTGCTTGTTCACATTTAAAGTTT
 CTCAACAGAACGGTGGAGAACATCTAGCTTATTITACAGTACATAAAG
 ATTGTGAGAACGAAGCTTCTTAAAGTTATGAAAATGTTAGATTAGCAG
 ATAATTGGACTGCTAATGTTCTGAGCAAGAAATATCAAGAGGCAATTGCTA
 ATATTAAAGGACAAATTAGACAAGGAAACTTATCAAGTAAATTATACA
 CTAGAGCTTAGCCAACAATTATGCTCGGATCC

MHIETVIDFKELGKRYRFKNPTKELIADTLEQVLEVIKEVDYYQSQNYVVGY
 LSYEASAAFDSHFKVSVQQKLAGEHLAYFTVHKDCENEAPLSYENVRLADNW
 TANVSEQEYQEAIANIKGQIRQGNTYQVNVTLELSQQLCSD

Sequence description:

- A] Length: 440 bp - 146 aa (partial gene sequence)
- B] This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-103 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-103 gene sequence.
 Shine Dalgarno sequence present upstream of
 ATG start codon, No apparent leader peptide sequence

ID-176

Clone 2-18/22b(b) (ID-104b)

GTGAATAATATGTTTATCTAAAATAGCCTGGCATAATTAAAACATTCT
 ATAGACCAGTACATACCATTCTCTTAGCCAGTTATTACTTTATTCAATTGA
 CTTGTTCTACGCTACTAATCTTAATGAGTGCTGTTGGAAAGAGATATGGGA
 CAGCGGCAACGGTTCTTCTGGAGTGATTGTTGTCAATCTTGCAGGT
 AGTCATGGAACATTATAGCTACAATATCTTGATGAAACAGCGTAGTAGTG

FIG. 1 CONT'D

AATTTGGACTGTATAACATTGGGGATGAATAAACGTCAAGTTGCGCGTG
TAGCTAGTCTAGAGCTGTTATTATTATATATTCTTATTCTATAGGAAG
TCTGTTAGTGCTTTTGTCAAATTATTAAATTGTCAACATTA
TTAACATCATGCACTAAATCTTAGTTAAGTTATGCCATTATTATTG
TATCGTTATTTACAGGTATTTGACTTAAAGTTAGAAAGTCCAGTTATCGA
CATGTTCATTTATCATCCCCATTAAGTCTTTAGAAAGAAACACAGGGA
GAAAAAGAACCAAAAGGTAACTTAACTTGCATTAGCGTTAGTAGCT
ATCGCCATCGCTTATACAATGGCTTACTTCAGGTAAAGCACCTGCATTA
GCTGTTATCTATCGTTCTTGCAGTACTTTAGTAATTGCTGGTACTT
ATCTTTTATATTAGTTATGACATGGTACTTAAAAGGTTGCGTCAAAA
CAAGCATTATTATAAACTGAGCATTGTATCAACTTCGCAAATGAT
TTTCGAATGAAGCAAAATGCAGTAGGGTTAGCAAGTATCACTTATTAGC
TGTTATGGCTCTAGTTACTATTGCTACAACAGTCTCACTCTATTCAAATACA
CAAATGTGTTACCGGACTATTCCAAAATCAGTAAGTTATCAATAGAT
AATTCAAAAGGTGACGCGAAAAATATATTGAAGAAAAGATTGAAAGAA
ACTAGGTAAAGTCATCAAGGAAGCTATCACTTATAATCAGACAATGATTG
GATGCCAGTTAGTCATCAAGTGACTIONTACACATCTA

MNNMFYLKIAWHNLKHSIDQYIPFLLASLLLYSLTCSTLLILMSAVGRDMGTA
ATVLFLGVIVLSIFAVVMEHYSYNILMKQRSSEFGYNILGMNKRQVARVASL
ELFIYIIFLISIGSLFAFFAKFIYLIFVNIINYHALNLSLSWPFIICIVIFTGIFTLE
VPVIRHVHLSSPLSLFRKKQQGEKEPKGNLILAILALVAIAIAYTMALTSGKAP
ALAVIYRFFFAVLLVIAGTYLFYISFMTWYLKRLRQNKHYYKSEHFVTSQM
IFRMKQNAVGLASITLLAVMALVTIATTVSLYSNTQNVVTGLFPKSVSLSIDNS
KGDAKNIFEKILKLGKSSKEITYNQTMISMPVSQSSDLISHL

Sequence description:

- A] Length: 1119 bp - 373 aa (partial gene sequence)
- B] This gene sequence was not identified using the LEEP system. It was identified upstream of the ID-104 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-104 gene sequence.
- Possible Shine Dalgarno sequence present upstream of a GTG start codon. Possesses a potential leader peptide sequence

ID-177

FIG. 1 CONT'D

Clone 2-5b (ID-112b)

ATGGTTGAGCCAATTATTCAATACAAGGACTTCATAAAAGTTTGGGAAA
 AATGAGGTTTAAAAGGCATTGACTGGATATTCAAGGAGAAGTGGT
 GGTTATTATTGCCCTCTGGCTCTGGTAAGTCACACATTAAAGAACAAAT
 GAATCTCTTGGAAAGTACCAACAAAGGGAACAGTGACTTTGAAGGGATTG
 ATATAACAGACAAAAAGAATGATATTAAATGCGCGAAAAAATGGC
 ATGGTTTCAACAGITCAATCTATTCCCAATATGACTGTACTAGAAAAT
 ATTACTTATCACCTATTAAGACAAAGGGACTTCTAAGCTTGTACTGCTCAG
 ACAAAAGCATACGAGCTACTGAAAAAGTGGACTCAAAGAGAAGGCTAA
 TGCTTATCCAGCAAGCTTATCTGGAGGACAACAAACACGGATTGCTATTG
 AAGAGGTCTTGCATGAATCCTGATGTCCTTGTGAACTACTTCA
 GCTCTGATCCTGAAATGGTAGGTGAAGTCTGACTGTTATGCAAGATT
 GCTAAATCTGGTATGACGATGGTTATTGTCACTCATGAAATGGGTTTGCA
 CGTGAAGTAGCGGATCGTGTCAATTATGGATGCAGGGATTATTGTTGAG
 CAAGGGACCCCTAAGAAAGTATTGAGCAGACAAAAGAAATCCGCACAAG
 AGACTTCTTAAGTAAAGTATTATAA

MVEPIISIQGLHKSFGKNEVLKGIDLDIHQGEVVVIIGPSGSGKSTFLRTMNLE
 VPTKGTVTFEGIDITDKKNDFKMREKMGMVQQFNLPNMTVLENITLSPIKT
 KGLSKLDAQTAKAYELLEKVGKEKANAYPASLSGGQQQRIAIARGLAMNPDV
 LLFDEPTSALDPEMVGEVLTVMQDLAKSGMTMVIVTHEMGFAREVADRVIF
 MDAGIIVEQGTPKKVFEQTKEIRTRDFLSKVL*

Sequence description:

- A] Length: 735 bp - 244 aa (full length gene)
- B] This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-112 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-112 gene sequence. Shine-Dalgarno sequence precedes the 'ATG' start codon. No obvious leader peptide

ID-178

Clone 2-5c (ID-112c)

FIG. 1 CONT'D

SUBSTITUTE SHEET (RULE 26)

ATGTCTCAsTATCAAGAGTGGTAGAAAACGACTCACTCGTAAAGATATT
 AAGTCAGATTAGAAGCTATAAAGGAGATGAATCTGAAATTCAAGGATCG
 TTTTACAAAACATTAGAATTGGAACGGCGGGATTGAGAGGTAAACTG
 GAGCAGGAACCAATCGTATGAATACTTATATGGTGGGGAAAGCAGCACAA
 GCATTAGCTAATCGATTATTGATCATGCCCTGAAGCTATTGCACGTGGAA
 TTGCAGTTAGTTATGATGTCCCCTATCAATCTAAGGAATTGCAGAATTA
 ACTTGGTCATTATGGCAGCAAATGGTATTAAAGCCTATTTA

MSHMNYKEIYQEWLENDSLGKDIKSDEAIKGDESEIQDRFYKTLEFGTAGLR
 GKLGAGTNRMNTYMGKAAQALANRLLIMALKLLHVELQLVMMMSRYQSKE
 FAELTWSIMAANGIKALYL

Sequence description:

- A] Length: 366 bp - 122 aa (partial gene sequence)
- B] This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-112 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-112 gene sequence. Shine-Dalgarno sequence preceded the 'ATG' start codon. No obvious potential leader peptide sequence.

ID-179

Clone 2-5d (ID-112d)

ATGCAACCTGTAAAAGTCGATGAACCTCTGTTGAAGAAAACCATTACTATT
 TTGAAAGGTATCCAAAAAAATACGAAGATTATCATCACGTAATATAA
 TAATGATGCCATAGAACGAGCTGCAGTACTATCTAATCGTTATATCCAAGA
 CCGCTTTTACCTGATAAAAGCAATAGACTTATTAGATGAAGCTGGTTCTAA
 AATGAACCTAACACTAAATTGTTGATCCAAAAGAAATTGATCAACGTCT
 CATTGAAGCAGAAAATTAAAGCGCAAGCGACTCGTGAAGAAGATTACG
 AACGTGCAGCTTACTTCCGTGACCAGATTGCAAAATATAAAGAAATGCAG
 CAACAAAAGGTGACGATCAAGATAACCTATTATTACGAAAAAACAAAT
 TGAGCACATCATTGAAGAAAAACGAATATCCCTGTTGGTGATTAAAAG
 AAAAAGAACAACTCAATTAAATTAAATCTCGCAGATGACTTGAAACAGCAT
 GTGATCGGCCAGGATGACGCTGTCATTAAGATTGCAAAAGCTATTGTCGT
 AATCGAGTTGGTCTTGGTAGCCAAACCGTCTATTGGTCCCTTTATTG
 TAGGACCAACCGGTGTTGGTAAAAGTGAACCTTCTAAACAATAGCAATTG
 AGCTCTTGGTTCAGCTGATAGTATGATTGCTTGTATGTCAAGAGTACAT
 GGAAAAGCATGCTGTTGCTAAATTAGTCGGAGCGCCTCCAGGATACTGG
 GATACGAGGAAGCTGGACAACTAACGTAAAAGGTTGTCGAAATCCCTAC
 TCGCTCATCCTCTAGATGAAAAGCTCATCCGATGTCATGCAT

FIG. 1 CONT'D

ATGTTCTGCAGGTCTTGATGACGGTCGATTAACAGATGGACAAGGAAG
 AACTGTTAGTTAAAGATACCATTATCATCATGACCTCAAATGCTGGTTC
 TGGTAAAACCTGAAGCAAGTGTGGCTTGGTGCCTCACGAGAAGGTAGGA
 CGAATTGAGCTCGGTACCCGGGATCCTCTAGAGTCGACCTGCAGGCAT
 GCAAGC

MQPVKVDPSVEETITLKGIQKKYEDYHHVKYNNDIEAAA VLSNRYIQDRF
 LPDKAIDLLEAGSKMNLTLNFVDPKEIDQRLIEAENLKAQATREEDYERAAY
 FRDQIAKYKEMQQQKVDDQDTPIIETKIEHIIEEKTNIPVGDLKEKEQLSQLINL
 ADDLKQHVIGQDDAVIKIAKAIRRNVRGLGSPNRPIGSFLFVGPTGVGKTELSK
 QLAIELFGSADSMIRFDMSEYMEKHAVA KLVGAPPGYVGYEEAGQLTEKVRR
 NPYSLLDEIEKAHPDVMHMFLQVLDDGRLTDGQGRTVSFKDIIIMTSNAGS
 GKTEASVFGASREGRTNSSSVPGDPLESTCRHAS

Sequence description:

- A] Length: 1070 bp > 356 aa (Partial gene sequence)
- B] This gene sequence was not identified using the LEEP system. It was identified upstream of the ID-112 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-112 gene sequence. Shine-Dalgarno sequence preceded the 'ATG' start codon. No obvious potential leader peptide sequence.

ID-180

Clone 2-7b (ID-113b)

ATGAGAGGAAAGGTATTACGGCACAACCCATTAGGTCTTTCTATTCTTTCTATTCTATTCTATTGGATTCTTAAGCATCACATCGAGAGAATA CATCATC ATCGTATAAACAGGTTAGATGCGAAGAGTGATTAAACAGGATTAAAACC CATTGCCATTATCAGCATTGATAACAAAGCAACAAGTTATTCTCTTGTAA CAAAAAGAAGGCGGAAATATGTCAAAGCTAGGGATAATATTAATGTTGATATCGAATTACGGGATTCTCCAAGTAGATCACATCATTTATCAGAAAAGCC GAGAATTAGGACAAAGGGTTAATATCATATAGAGGAAATTCTCTCGTT ACTTGATAAGAAGTCATTGAAAGTTAAGTTACTAATAAGTTAAAGGAAAAGAAGCATCGATTAGCAGGAATGCCTAAAGAATCGGAGTGGGTATTG CATGGTCCCTTCTAGACAGAACATTATTAAGAAATTATCTGAGTTATAAT

FIG. 1 CONT'D

ATTGCTGGTGAGATTATGCCTATGCCCAAACGTTCGCTACTGTGAGTTAT
TTGTCAATGGTGAGTATCAGGGAG

MRGKVIYGTTLIGLFLFLFFYWIPKHHIERIHHRIKVQDAKSDLTGFKTHLPII
SIDTKQQVPLVTKEGGKYVKARDNINVDELRDSPSRSHLSEKPRI RTKGLIS
YRGNSSRYFDKKSLVKVFVTNKLKEKKHRLAGMPKESEWVLHGPFLDRTLLR
NYLSYNIAGEIMPMPQTATVSYLSMVSIRE

Sequence description:

- A] Length: 582 bp - 194 aa (Partial gene sequence)
- B] This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-113 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-113 gene sequence.
- ATG start codon is preceded by a Shine-Dalgarno sequence-Possesses a potential leader peptide sequence. C-terminus to be determined.

ID-181

Clone 2-17b (ID-117b)

CTTCACATTTATTGATCACTATCTGACAAATGTTAATCAAACAGCAGTTCT
TATTTAGTGGATATTATTCAATGTATGTCAGACCTTAATTCAATAT
TTGGGAATCTCTTTGCGCGTGTCTTATAGTATTGTTAGAGATATTC
GTAGAGATGCTTGCTAATATGAAAGGCTAGGCATGTCTATTGATA
GGACACCGGCAGGATCTATTGTCACGTATTACTAATGATACTGAAGCAA
TATCTGATATGTTTCGGGTATTTATCAAGTTATCTCGGCGATATTAT
TTTACAGTTACTCTGTACACTATGTTGATGCTAGACATTAAACTAACAGG
ACTCGTCGCTTTGTTACCTGTTATCTTATATTAGTGAATGTCTATCGG
AAAAAATCAGTCAGTCACTGTCTTAAACGAGAAGTTACTTAGTGAATTC
AACAGTAAATTATCAGAAAGTATTGAAGGAATT

SHFIDHYLTNVNQTAVLILVGYYSMYVLQLIYFGNLFFARVSY SIVRDIRD
AFANMERLGM SYFDRTPAGSIVSRITNDTEAISDMFSGILSSFISAIFIFTVTLYT
MLMLDIKLTGLVALLPVFILVN VYRKKS VTVIAKTRSL SDINSKL SESIEGI

FIG. 1 CONT'D

Sequence description:

- A] Length: 498 bp - 165 aa (Partial gene sequence)
 B] This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-117 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-117 gene sequence. N- and C-termini have yet to be determined

ID-182**Clone 3-8b (ID-120b)**

ATGTACCATATTGAATTAAGGAAAGCTTACTACCAAGAGAACGCCT
 AGTTGATTAGGCGCAGATAGATTGAGTAATCAGGAGTTATTAGCCATTCT
 CTTACGTACAGGTATTAAAGAAAAACCTGTTCTGAAATTCAACGCAAAT
 TTTAGAAAACATAAGCAGTTAGCAGATTGGTCAATTATCCTTACAGGA
 GTTGCATCCATTAAAGGAATCGGTCAAGGTTAAATCCGTCGAAATAAAAG
 CTATGCTAGAACTAGCAAAACGGATTACAAAGCTGAATATGATCGTAAA
 GAGCAAATTAAAGTAGTGAACAATTAGCGAGGAAATGATGCTCGAATT
 AGGGGATAAAAACAAGAACATTAGTAGCTATTATATGGATACACAAA
 ATCGTATTATCGAACAGAGAACTATTATTGGTACTGTACGTCGTTCA
 TAGCAGAGCCAAGAGAAATTCTACATTATGCTTGTAAAAACATGGCAACT
 TCTTGATTATTATACATAATCATCCCTCAGGTTCTCAAATCCCAGTGAAA
 GTGATTAAAGTTCACTAAAAAAATAACGATCATGTGATCATCTGGGAA
 TTGTCTGCCTAGATCACATCATCGTGGAAAAATAATATTAGTTTC
 GAGAAGAAGCAGATATTATATAA

MYHIELKKEALLPRERLVDLGADRLSNQELLAILLRTGIKEKPVLEISTQILENI
 SSLADFGQLSLQELQSIKGIGQVKSVIEKAMLELAKRIHKAEYDRKEQILSEQ
 LARKMMLELGDKKQEHLVAIYMDTQNRIIEQRTIFIGTVRRSVAEPREILHYAC
 KNMATSLLIIHNHPSPNPSESDSLFTKKIKRSCDHLGIVCLDHIVGKNKYYSF
 REEADIL*

Sequence description:

- A] Length: 681 bp - 227 aa (full-length gene)
 B] This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-120 gene which was identified by LEEP,

FIG. 1 CONT'D

during cloning and sequence analysis of the full-length ID-120 gene sequence.
 ATG start codon is preceded by an typical
 Shine-Dalgarno sequence. No obvious leader
 peptide sequence

ID-183**Clone 3-11b (ID-121b)**

```
TGGTTAAAAGTAGTGTAGCTTGTATTCCATCTATTTAATTGCTTACCAT
TTGATAATTGGTTGAAGCTCATTAAATTCAATGATTCCGATTGCAATAGC
CCTAACATCTTATGGTTTGTCATATGGGTTGAAAAACGTAATGCACAC
CTCAAACCACAGGTAAACCGAATTGGCAAGTATGTCTACAAGACAGCTTC
TTGATTGGATGTTCCAGGTTCTCAGTATTGTCGGGAACCAAGTCGTTCTG
GAGCTACTATTAGGAGCAATTATTATTGAACTAGTCGTTCGGTGCTG
CTGACTTACCTTCTCCTGCCATCCCAACTATGTTGGTTAGTGGACT
TAAGGCAGTTAAATATTAGATGGTAACGTCTTGAGTTAGACCAATC
TTAATACCTTATAGCAAGTCTGACAGCTTCGTAGTTAGTTATATGTT
ATTGTTCTGACAGACTATGTCAAACGACACGATTCAACCATTGGT
AAGTATCGTATAGTCTAGGAAGTTACTCATCCTCTACTGGTTAGTTGTT
ATTATTCTAA
```

```
WLKVVIACIPSILIALPFDNWFEAHFNFMPIALIIFYGFVFIWVEKRNAHLKP
QVTELASMSYKTAFLIGCFQVLSIVPGTSRSGATILGAIIGTSRSVAADFTFLA
IPTMFGYSGLKAVKYFLDGNVLSDLQSLILLVASLTAFFVSLYVIRFLDYVKR
HDFTIFGKYRIVLGSLLILYWLVVHLF*
```

Sequence description:

- A] Length: 579 bp - 193 aa (partial sequence)
- B] This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-68 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-68 gene sequence described in WO 00/06736. N-terminus has yet to be determined.

ID-184**Clone 3-11c (ID-121c)****FIG. 1**_{CONT'D}

ATGGAAATGAAACAAATCACTGAAACAAACTGAAAATTACAATTAGTAT
 GGAAGATTAGAAGATCGGGTATGGAGCTGAAAGATTCCTAACCCCTCA
 GGAGAAGACTGAGGAATTTCTATTCTCATGGATGAATTAGACTGCC
 AGAAAACCTAAAAATAGTGGTATGTTAAGTTTCGAGTAACACCTAAAA
 AAGATCGCATTGATGTTTGTACAAAGTCTGAATTAAGTAAAGATTAA
 ATTAGAAGAATTAGCAGATTGGGTGACATTCAAAAATGTCTCCAGAAG
 ACTTTTTAAAACCTTGGAACATCGATGTTGGAAAAGGGGATACGGAT
 GCCCATGCCAAATTAGCAGAAATTGAAAATATGATGGATAAGCAACTCA
 AGAAGTAGTTGAGGAAAATGTTCTGAAGAACACCTGAAAAGGAAGTAG
 AAACGATTGGATATGTTCACTATGTCCTTGATTTGATAATTGAAGCTGT
 AGTCGATTTCACAAACGATTGATTTCCAATAGAAGCTT

MEMKQISETTLKITISMEDLEDRGMELKDFLIPQEKTTEFFYSVMDELDLPENF
 KNSGMLSFRVTPKKDRIDVFVTKSELSKDLNLEELADLGDISKMSPEDFFKTLE
 QSMLEKGDTDAHAKLAEIENMMDKATQEVEENVSEEQPEKEVETIGYVHY
 VFDFDNIEAVVRFSQTIDFPIEA

Sequence description:

- A] Length: 547 bp - 182 aa (Partial sequence)
- B] This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-68 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-68 gene sequence.
- ATG start codon is preceded by an typical Shine-Dalgarno sequence. No obvious potential leader peptide sequence

ID-185

Clone 3-16b (ID-122b)

GGAAACCAACGCCAGTACAATCGTCAAGGGTAGATTATCCTAAACGTAG
 TCGTGCAGATTGTAGAAGTTATTTAGACAAGCTCTACTACTGATTA
 TTCTGGTGTACAAAGGTTACTATATTGACTTGAAGCCAAAGAAACCCG
 GCAGAAAATGCTATGCCTATGAAAATTTCATGCTCACCAAATAGAGC
 ACATGGCAAATGTATTACAGCAAAAGGGATTGCTTGTCTGCTTCATT

FIG. 1 CONT'D

TTTCCACACTTAAGGAAACCTATCTACTCCCTGCTAATGAGTTAATTCAATT
TTATCAGATTGATAAAGGCAATAATCAATGCCTATTGATTATATCAGAAA
AAATGGATTTTCGTAAAGGAGAGTGCCTTCTCAAGTCCCTACTTAGA
TATTATTGAAGAAAAATTATTAGGCGGTGATTACAATTAA

GNQRPVQSSRVDYPKRSRAKIVEVYFRQASTTDYSGVYKGYYIDFEAKETRQ
KTAMPKNFHAHQIEHMANVLQQKGICFVLLHFSTLKETYLLPANELISFYQI
DKGNKSMPIDYIRKNGFFVKEAFPQVPYLDIIEEKLLGGDYN*

Sequence description:

A] Length: 447 bp - 149 aa (partial sequence)

B] This gene sequence was not identified using the LEEP system. It was identified upstream of the ID-122 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-122 gene sequence. N-terminus has yet to be determined

ID-186

Clone 3-17b (ID-123b)

GGATCCTAAAAACGCTAAGGTTATCAAAAAAATGCTGATCAATTAGTG
ACAAGGCAATGGCTATTGCAGAGAAGTATAAGCCAAATTAAAGCTGCA
AAGTCTAAATACTTGTGACTTCACATACAGCATTCTCATACTTAGCTAAG
CGATACGGATTGACTCAGTTAGGTATTGCAGGTGTCTCAACCGAGCAAGA
ACCTAGTCTAAAAAATTAGCCGAATTCAAGGAGTTGTGAAAACATATA
AGGTTAAGACTATTTTGTGAAGAAGGAGTCTCACCTAAATTAGCTCAAG
CAGTAGCTTCAGCTACTCGAGTTAAAATTGCAAGTTAAGTCCTTAGAAG
CAGTCCCCAAAACAATAAAGATTACTTAGAAAATTGGAAACTAATCTTA
AGGTACTTGTCAAATCGTTAAATCAATAG

DPKNAKVYQKNADQFSKAMAIKEKYKPKFKAAKS KYFVTSHAFSYLAKR
YGLTQLGIAGVSTEQEPESAKKLAEIQEFVKTYKVKTIFVEEGVSPKLAQAVAS
ATRVKIASLSPLEAVPKNNKDYLENLETNLKVLVKSLNQ*

Sequence description:

FIG. 1 CONT'D

A] Length: 433 bp - 144 aa (partial sequence)

B] This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-123 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-123 gene sequence. N-terminus has yet to be determined

ID-187

Clone 3-46/47 (ID-130b)

```

ATGAAAAAAGTCATCGATTAAAAAAACTACAAAAAGCATACGCCTCAGA
AACTGTTTAAATAATTAATTGGAGGTGTTAAAGGGAGAAATAATTGG
ATTAATAGGACCCTCTGGAGCAGGGAAATCTACCTTGATTAAAATATGCT
TGGCATGGAAAAAGCAGATAAGGGAACAGCTCTTGTCTTGATACTCAAA
TGCCAGATCGTAATATTTAAATCAAATTGGCTATATGGCTCAATCTGATG
CCTTACACGGAGTCTTAACTGGCTTAGAAAATTTATTATTCTTGGAAAAAA
TGAAAGGTATTCAAAAAACTGAATTAAAACAGCAGATAACTCATATTCT
AAAGTAGTAGATCTAGAAAACCAACTTGATAAAATTGTCTCAGGTTACTCA
GAAGGTATGAAAAGACGGCTTCTAGCCATGCCCTACTTGGAAACCCC
ACAGTTTAATCCTAGATGAACCTACCGTTGGAATTGATCCATCCTTGAGG
AGAAAAATCTGGCAAGAGCTAATTAATTAAAGGATGAAGGACGTTCTAT
CTTATTACAACCCACGTTATGGATGAAGCAGAATTACAAGTAAGGTTGC
ACTACTATTACGTGGAAACATTATTGCCTTGATACTCCATTACATTAAA
AAAACAATTAAATGTGAGTACTATTGAGGAAGTTTCTTAAAGCTGAAGG
AGAATAA

```

```

MKKVIDLKKLQKAYASETVNNINLEVFKGEIIGLIGPSGAGKSTLIKTMLGME
KADKGTLVLDTQMPDRNILNQIGYMAQSDALHESLTGLENLFFGKMKGIQ
KTELKQQITHISKVVVDLENQLDKFVSGYSEGMKRRLSLAIALLGNPTVLILDEP
TVGIDPSLRRKIWQELINKDEGRSIFITTHVMDEAEELTSKVALLRGNIIFDTP
LHLKKQFNVSTIEEVFLKAEGE*

```

Sequence description:

A] Length: 717 bp - 239 aa (Possible full-length sequence)

B] This gene sequence was not identified using the LEEP system. It was identified upstream of the ID-130 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-130 gene sequence. ATG start codon is preceded by a possible

FIG. 1 CONT'D

Shine-Dalgarno. No obvious potential leader peptide sequence

ID-188

Clone 3-83b (ID-144b)

ATGGTACAAATGATACTGATATGATTAAAACAATTGAGCATTTGCTGAG
 ACACAAAGCTGATTTCCAGTGTATGATATTTAGGGGAAGTCCTACTTAT
 GGACAACTAAAGTAGACTCTGACTCTAGCTGCTCATATTGATAGCCTA
 GCCCTTGTGAAAAATCACCTGTCTAGTATTGGTGGTCAAGAATATGAA
 ATGTTGGCGACATTGTTGCTTAACAAAGTCAGGGCATGCTTATATACCG
 GTTGACCAACACTCTGCTTGGATAGAATAACAGGCTATTATGACAGTTGCT
 CAACCAAGCCTATCATTTCAATTGGTGAATTCTCTTGAAGTTGATAAT
 GTCCCAATCCTAGACGTCTCAAGTTCTAGCTATTGGTAAAGGTTGATAAT
 CCTTATGAGGTAAACACATTCTGTTAAAGGTGATGATAATTACTATATTATT
 TTCACTTCAGGGACTACTGGTTACCAAAAGGTGTGCAAATTTCACATGAC
 AATTATTGAGCTTACAAATTGGATGATTCTGATGATGAGTTTCAGTTC
 CTGAAAGACCGCAAATGTTGGCTAACCC

MVQMIHDMIKTIEHFAETQADFPVYDILGEVHTYGQLKVDSDSLAAHIDS LGL
 VEKSPVLVFGGQEYEMLATFVALTKSGHAYIPVDQHSALDRQAIMTVAQPSL
 IISIGEFPLEVDNVPILDVSQVSAIFEEKTPYEVTHSVKGDDNYYIIFTSGTTGLP
 KGVQISHDNLLSFTNWMSDDEFSVPERPQMLAQ

Sequence description:

- A] Length: 592 bp - 197 aa (partial sequence)
- B] This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-144 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-144 gene sequence.
- Putative ATG start codon is preceded by a typical Shine-Dalgarno sequence. No obvious leader peptide sequence
- This orf is not in frame with nuc

ID-189

FIG. 1 CONT'D

Clone 3-86b (ID-145b)

ATGGAAAATCATCGTTATGAAGATGAAGGTAAATTCCAGCGTAAGATGAC
 CAGTCGTACATCTCTTATGTTATCGCTAGGTGGTGTATCAGGGACTGGGCTT
 TTCTTGAGTCAGGTTATACCAATTGCACAGGCTGGTCCGCTGGAGCTGTG
 CTGTCTTATTGATTGGTGCCATGCCGGTACGGGGCATCCACACTTATGCCA
 CTAAGTTATCAGTCCTGGAACAGGTTTACTGTTGCTGGCTATATTGGAT
 TTGTTGGACGGTCGCCCTGGGGACTGAATTTAGGTGCTGCCATGCTGAT
 GCAGCGCTGGITCCCAAATGTGCCGGCTGGGCATITGCTTCCTTTGCC
 CTTGTGATTTGGTTAAATGCTCTAGCGTAGCCTTTGCAAGAACAG
 AGTCCTTCTCAAGTATTAAGGTTATGCTATCATTATCTTATTATCTG
 GGCTTAGGTGCTATGTTGGCTAGTTCCCTTGAAGGTACAGCACAAAGGCT
 ATTCTCTTCACTCATGACTGCCAATGGTGCCTTCCAAATGGTATCGTTG
 CAGTTGTCTCAGTCATGTTGGCTGTTAACTATGCCCTCTGGTACTGAGTT
 AATTGGTATTGCGGCTGGTAAACGGATAATCCAAAGAACAGTGTACCAA
 GGGCTATTAAAACGACAATCGGTGCGCTGGTTTTCTTGACTGACAA
 TTGTTGTCCTAGCTTCGCTATTGCAAATGAAAGAGGCAGGCGTATCCACAG
 CACCATTGTTGATGCTTTGACAAGATGGGAATCCCTTACGGCGGATA
 TCATGAACTCGTTATCTTGACAGCCATCTGTCTGGTAACTCAGGTCT
 CTACGCATCAAGCCGTATGCTCTGGTCCCTGCCAATGAAGGTATGTTGTC
 AAAATCTGTTGTAAAATCAATAAACACGGTGTCCAATGCGTGTCTTCT
 CTTGTCAATGGCAGGAGCAGTGCTGCGCTTTCAAGTATTACGCTGC
 AGACACAGTTATCTAGCCTGGTTCAATCGCGGGCTTGCTGTTGTC
 GTATGGCTAGCCATTCAGTCGCACAAATCAATTCCGCAAGGAATT

MENHRYEDEGKFQRKMTSRHLFMLSLLGGVIGTGLFLSSGYTIAQAGPLGAVL
 SYLIGAVVVLVMSLGELEVAMPVTGSFHTYATKFISPGTGFVAWLYWIC
 WTVALGTEFLGAAMLMQRWFNPVPAWAFASFFALVIFGLNALSVRFFAEAES
 FFSSIKVIAIIIFIILGLGAMFGLVSFEGQHKAILFTHTANGAFPNGIVAVVSVM
 LAVNYAFSGTELIGIAAGETDNPKEAVPRAJKTTIGRLVVFFVLTIVVLASLLPM
 KEAGVSTAPFVDVFDMKGIPFTADIMNFVILTAILSAGNSGLYASSRMLWSLA
 NEGMLSKSVVKINKHGVPMRALLSMAGAVLSLFSSIYAADTVYLALVSIAGF
 AVVVVWLAIPVAQINFRKEF

Sequence description:

A] Length: 1126 bp - 393 aa (partial gene)

FIG. 1 CONT'D

sequence)

B] This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-145 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-145 gene sequence. Putative ATG start codon is preceded by a typical Shine-Dalgarno sequence. Possesses a possible leader peptide sequence.

ID-190

Clone 3-94b

TCAGAAAATGCAGAGGCAGCAACGGTTGCCACAAACTGGTTACCAAAGG
 AGCTAATGTCATTATCGGACCAGCAACATCGGGTGCAGCTGCATCTTCAAC
 TCCAAAAGTAATGCAGCAGCAGTTCCAATGATTGCACCTGCTGCGACAC
 AAGACAATTAGTCTATGGTTCTGATGGAAAAACCTAAATCAGTATTCT
 TCCGAGCTACTTTGTCGATAATTATCAAGGAAAGCTATTGTCTCAGTATG
 CTACAGACAACCTAAAGCTAAAAAAGTTGTTCTATTATGATAATTCAT
 CAGATTACTCAAAGGGGTAGCAAATCATTAAGGAAAGTTAGTGGA
 AAAATTGTTGATAGTATGACATTCTCCGCTGGTGATACTGATTCCAAGCG
 TCATTGACTAAGTTGAAAGGGAAAGAATATGATGCTATTGTGATGCCAGG
 TTACTATACCGAGACAGGATTAATAGTTAAGCAAGCGCGTGATTAGGTAT
 CTCTAAACCGGTTCTGGGCTGATGGTTTGATAGTCCGAAATTGTGCA
 ATCGGCAACACCTGTAGGAGCTCAAACGTTATTATTGACAGGTTTCAC
 TACACAAGGATCAACCAAAGCTAAAGCT

SENAEAATVATNLVTKGANVIIGPATSGAAASSTPKVNAAAVPMIAPAATQD
 NLVYGSdgKTLNQYFFRATFVDNYQGKLLSQYATDNLKAKKVVLFYDNSSD
 YSKGVAKSFKEYSYGKIVDSMTFSAGDTDFQASLTKLKGKEYDAIVMPGYYT
 ETGLIVKQARDLGISKPVLDGFDSPKFVQSATPGASNVYYLTGFTTQGST
 KAKA

Sequence description

A] Length: 637 bp - 231 aa (partial sequence)

FIG. 1 CONT'D

SUBSTITUTE SHEET (RULE 26)

92 / 110

B] This gene sequence was not identified using the LEEP system. It was identified upstream of the ID-149 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-149 gene sequence. N- and C-termini have yet to be determined

ID-191

Clone 2-c94b (ID-153b)

TTGGGACTTAAAGACCATGCTTAGTCTATCCATTTCATTATCTGGGGGG
CAAAAGAACGTGTCGCACTAGCTCGCGATGATGATTGATCCACAGATT
ATTGGTTATGATGAGCCAAGTAGCGCTTGATCCAGAGTTGCGTCAAGAA
GTAGAAAAACTAATTTCACAAAATAGAGAAACAGGTATGACACAAATTGT
AGTAACACATGATCTCAATTGCTGAAAGTATCTGATACGATTCTCAA
AATTAATCCTAAGTAG

MGLKDHALVYPFSLSGGQKQRVALARAMMIDPQIIGYDEPTSALDPELRQEV
EKLILQNRETGMTQIVVTHDLQFAESISDTILKINPK*

Sequence description

A] Length: 270 bp - 90 aa (partial sequence)

B] This gene sequence was not identified using the LEEP system. It was identified upstream of the ID-153 gene which was identified by LEEP, during cloning and sequence analysis of the ID-153 gene sequence.
N-terminus has yet to be determined

ID-192

Clone 2-c1b (ID-155b)

ATGACTAATATCTCAGATGTTCCAAAAGCTATTAGAACACAGGCACAGTAT
GTTCTCTGGGAATGAGAGTTATGGATCAGTCGGTATTACCGAAAACATAT
AATTCAAAAGAACCTTATTGAAACCAGATATGATTATTCATGATAGA

FIG. 1 CONT'D

AGACAAGAGACAATGCTAAAATCACTCAAGAAATAGAAATGGAGCATTG
A

MTNISDVPKAIRTQAQYVLLGMRVMDQSVLPKTYSKEPYLKPDIMIYIHDRR
QETMLKITQEIMEH*

Sequence description

A] Length: 204 bp - 68 aa (partial sequence)

B] This gene sequence was not identified using the LEEP system. It was identified upstream of the ID-155 gene which was identified by LEEP, during cloning and sequence analysis of the ID-155 gene sequence.

ATG start codon is preceded by a potential typical Shine-Dalgarno sequence.
Has a
typical leader peptide. N-terminus has yet to be
verified

ID-193

Clone 2-54altb (ID-172b)

AAGCTTGATGCCTGCAGGTCACTCTAGAGGATCTGGGAATAAATT
TGGATTTCATGACGATGTAAAGCCAATTATTCAGGGAAAGGTCTAAA
TGAGGCTTTATCGTGAGTTATCTGCAGCTAAGGGTGAACCTGAGTGGAT
GTTGGACTTCGTCTAAATCCTGGAAACGTTAATAAAATGCCGATGCA
GACCTGGGGAGCAGATTATCAGATATTGATTTGATGATATTATTTATA
TCAAAAAGCATCTGATAAACCTGCGCGTGAATGGATGATGTTCCAGAAA
AAATCAAAGAAACTTTGAAAGAATTGGGATCCAGAAGCTGAAAGAGCC
TATCTGCAGGAGCATCAGCACAATATGAATCAGAAGTAGTTATCACAAT
ATGAAAGAAGAATATGATAAGCTGGGTATTGTTTACGGATACTGACTCT
GCACTTAAAGAGTACCCAGAGCTATTCAAAAAATATTTGCTAAACTTGTCT
CCTCCAACAGATAATAAATTAGCTGCTCTGAACCTGCTGTATGGTCAGGT
GGAACATTATTTATGTCCTAAAGGTGTTAAGGTGGATATTCCACTTCAA
ACTTACTTCCGTATTAATAATGAAAATACTGGACAATTGAACGTACTCTC
ATTATTGTTGATGAGGGAGCAAGTGGTCACTATGTTGAAGGTGTACCGCC
CCAACTTATTCTCAAATAGTTACATGCAGCTATAGTTGAAATTTCAC
TTGATGGAGCTTATATGCGCTATACGACTATTCAAAATTGGTCCGATAATG
TCTATAATTAGTGACAAAACGTGCTACCGCTAAAAAGATGCAACAGTT
GAGTGGATAGATGGAAATCTAGGAGCTAAAACAACAATGAAATACCCATC

FIG. 1 CONT'D

GGTTTACCTTGATGGTGAAGGAGCACGTGGCACGATGTTGCTATTGCTTT
 TGCAAACAAAGGACAACACCAAGATAACGGGTGCAAAGATGATTCTATAATG
 CCCCCCATACTAGITCATCCATTGTCCTAAATCAATTGCTAAGGGTGGGG
 GAAAAGTTGATTATCGAGGTCAAGTGACATTTAATAAAAGATTCCAAAAAA
 TCAGTGTACATATAGAATGTGACACCATTGATGGATGATATTCAAAAA
 TCAGATACCATACCGTTAACATGAGATTCTATAATTACAGGGTGTAGAG
 CATGAAGCAAAGGTGCTAAGATTCTGAAGAGCAACTGTACTACTGATG
 AGTCGAGGTTATCTGAAGCTGAAGCAACAGAAATGATTGTTATGGGGTT
 GTTGGCCCTTACGAAAGAATTACCAATGGAATATGCGGTAGAGTTAAA
 TCGTTAACCTGAAATGGAAGGTTAGTTAA

MHACRSTLEDLGEYKFGFHDDVKPIYSTGKGLNEAVIRELSAAKGEPEWMLD
 FRLKSLETFNKMMPMQTWGADLSIDFDDIYYQKASDKPARDWDDVPEKIKE
 TFERIGIPEAERA YLAGASAQYESEVVYHNMKEEYDKLGIVFTDTDSALKEYP
 ELFKKYFAKLVPPTDNKLAALNSAVWSGGTFIYVPKGKVVDIPLQTYFRINNE
 NTGQFERTLIIVDEGASVHYVEGCTAPTYSSNSLHAAIVEIFALDGAYMR YTTI
 QNWSDNVYNLVTKRATAKKDATVEWIDGNLGAKTTMKYPSVYLDGEGARG
 TMLSIAFANKQHQDTGAKMIHNAPHSSIVSKSIAKGGKVDYRGQVTFN
 KDSKKSVSHECDTILMDDISKSDTIPFNEIHNSQVALEHEAKVSKISEEQLYYL
 MSRGLSEAEATEMIVMGFVEPFTKELPMYEAVELNRLISYEMEGSVG*

Sequence description:

- A] Length: 1411 bp - 469 aa (Possible full-length gene)
- B] This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-72 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-72 gene sequence. No obvious Shine Dalgarno sequence upstream of TTG start codon insufficient sequence data). N terminus needs verification.

ID-194

Clone 3-1b (ID-81b)

ATGATAGAATTCTTCTAATATCAGAACAGAGATTCCGCAGATGCCTTA
 CTTATCCATAGTTGATTTATCTGTCTTACCTTCTGATGTGGCTGACTTT
 GGTTAACAGAGATAAGCCTTGTATAAAACTATTGGAGTATCCTTTAGG
 ACTTCAGTTAACAGATTATACTTGGTTCTGGCAAAATTGCCTTA

FIG. 1 CONT'D

TCTGAAAGTCTTCCCCTTACCATGGTCGAATAGGCATGTTGTCGGTCTCT
TA

MIEFFSNIRTEIPQMPLLHSLISLVLPMWLTLVNRDKPLYKTIWSILLGLQLI
TIYTWFYWAKLPLSESLPLYHCRIGMFVGLL

Sequence description

- A) Length: 261 bp - 87 aa (partial gene sequence)
- B) This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-81 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-81 gene sequence. Sequence Characteristics: Possesses a potential leader peptide sequence Orf is preceded by a potential Shine-Dalgarno sequence.

ID-195

Clone RS-55b

AAGCTTGTGCAAAGTATTAAAGAGATAGGATTAGCTAATGCGCATTATTAA
GCTGTTGCTCCGACAGGGTCAATCAGTTATCTTCTTCTGTACTCCGAGCC
TTCAACC GGTTGTATCACCTGTCGAAGTACGCAAGGAAGGAGCACTGGGG
AGGGTTTATGTAGCTGCTTATAAGATTGATGCGAGATAATTATGTCTACTAC
AAAAAAAGGAGCTTATGAAGTGGGATCTGAGGGGATTATCAATATTGCAGC
TGCCGCTCAAAACACATTGATCAAGCTATTCGTTAACGCTTTCATGAC
AGATCAAGCAACTACCGCGAGATTAAATAAGCCTATATTCAAGCATITA
AACAAAAATGTGCCTCTATTATTATGTACGAGTGAGACAGGACATCCTAG
AAGGTAGCGAGAGTTATGATGATATGCTGGATGATTCACTTCATCGGACT
TAGAAGACTGTCAATCCTGCATGATTAA

>KLVQSIKEIGLANAHILLA VAPTGSISYLSCTPSLQPVVSPVEVRKEGALGRV
YVAA YKIDADNYVYYKKGAYEVGSEAIINIAAAAQKHIDQAISLTLMFTDQAT
TRDLNKAYIQAFKQKCASIYYVRVRQDILEGSESYDDMLDDFTSSDLEDQSC
MI*

Sequence description:

FIG. 1 CONT'D

A] Length 486 bp - 162 aa (Partial sequence)

B] This gene sequence was not identified using the LEEP system. It was identified upstream of the ID-87 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-87 gene sequence. N-terminus to be determined.

ID-196

Clone RS-59(ID-90b)

GTGAGGACATATATTACAAACTTGAATGGACATTCAATCACTAGTACAGC
 ACAAAATAGCTAAAACATGGTAACAGATATAGCAGTAAGCTTAGGTTTC
 GTGAGCTGGGAATACATTCTATCCGATTGATACTGATTCTCCTGAGGAAA
 TGAGTAAGCGTTAGATGGAATCTGTTCCGGACTTAGAAAAAAATGATATTG
 TCATATTTCAGACACCTACATGGAACACTACAACCTTTGATGAAAAATTAT
 TTCACAAATTAAAAATATTGGTGAAAGATTGTTATTTTATACATGATGTT
 TGTACCGCTAATGTTGATGGAAATTTTATTGATGGATAGAACTATAGC
 TTATTATAATGAAGCAGATGTTAATAGCCCCTAGTCAAGCAATGGTCGAT
 AAGCTT

MRTYITNLNGHSITSTAQIAQNMVTDIAVSLGFRELGHSYPIDTDSPEEMSKRL
 DGICSLRKNDIVIFQTPTWNTTFDEKLFHKLKIFGVKIVIFIHDVVPLMFDGN
 FYLMDRTIAYYNEADVLIAAPSQAMVDKL

Sequence description:

A] Length: 414 bp - 138 aa(partial gene)

B] This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-90 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-90 gene sequence.

No obvious signal peptide, but a possible Shine Dalgarno sequence is present upstream of ATG start codon. C-terminus has yet to be determined.

ID-197

Clone RS-59c (ID-90c)

FIG. 1 CONT'D

CATGGAAATGAAGTTGATGATGTTATTAGAAGGGCATTGAATATAATCAC
 CTTATCTTGCTTTGATAATACCTGTCTAACAGAGAGTTAGTATTAGATA
 GCAATATCATTCTCACACAACCTGTGAACAATTGATAAAATTAAATGAAAAA
 ATTATCAGGCTCCATTATGTATTGCTAGAGCAACAAAGAGAACAAACA
 AGTAATGAAACAAAAGAGCGTTATAAAGAAATTAGGAGGGTATGGAA
 ATGCCTAA

HGNEVDDVIRRAFEYNHLIFAFDNTCHNRELVLDNSIISHTCEQLINLMKNLS
 GSIMYLLEQQREQTSNETKERYKEILGGYGNA*

Sequence description:

- A] Length: 261 bp - 87 aa(partial gene sequence)
- B] This gene sequence was not identified using the LEEP system. It was identified upstream of the ID-90 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-90 gene sequence. N-terminus has yet to be determined

ID-198

Clone RS-70b (ID-93b)

ACATTTTATATTATGTATTGAAGACGTAGCCACCCAGTCAAATATGACT
 GGGAAAGATTTTAGTATGTCTAAAGAAGAGTTGTATATTACCCGTATT
 AAAACTTTTAAGAATCAAGGTGTATACAACCGCTGATTGGTCTATTCTC
 CTTATGGGTATATATTACAGAATCAAGAAATTGTAGCTATTITTA
 TCAATGTGTTGCTAGTTGCTTTATGGTGTCTTGACAGTTGATAAAAAAA
 TCTTATTAAAAACAGGGTGGTTACCTATATTAGCTCTTTAACATTCTATT
 TTAA

TFLYYVFEDVATQSNMTGKIFSMSKEELSYLPVIKLFKNQGVYNGLIGLFLY
 GLYISQNQEIVAIFLINVLLAVYGALTVDKKILLKQGGLPILALLTFLF*

Sequence description:

- A] Length: 312 bp - 104 aa (partial gene sequence)

FIG. 1 CONT'D

98 / 110

B] This gene sequence was not identified using the LEEP system. It was identified upstream of the ID-93 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-93 gene sequence.

N-terminus has yet to be determined

ID-199

Clone RS-70c (ID-93c)

ATGAAATTAAAGTGCCTTGATTATGGCCTTATTGATTATGGAAAAACTGCA
AGTGATGCAATACAAGAACGATTCTTTATCACAAAGAGGCGGAGCAACT
AGGCTATCATCAATTGGGTGGCTGAACATCACGGTGTAAAGGCATTCAAG
TATTAGCAATCCAGAATTAATGATAATGCATTGGCTAACCAAGACTAAATC
TATCAAAATTGGCTCTGGAGGTATAATGCCCTCTGCACTATAGTAGTTAA
ACTCGCGGAGACTCTCAAGACATTAGAGACATGTCATCCGAATCGAGTAA
GTATTGGTTAGGAAATTCACTAGGGACAGTTAAAGTTCAAATGCACTTC
GTAGCTTACATAAAGCACATGATTACGAAGAGGGTACTGGAGGAATTGAAG
TCATGGCTTATTGATGAATCATCCAGTAAGGAACCATTAGTTCAACCGACT
CTTCTAGCTTCCCAGACTTATATGTGTTGGGAGTGGTCAAAAATCAGCT
TATTTAGCGGCTAAACTGGCTTAGGCTTACCTCGGTGTTTCCTTTA
TGGACAAAGACCCATTGACAGAAAGCTAAA

MKLSVLDYGLIDYGKTASDAIQETILLSQEAEQLGYHQFWVAEHHGVKA
FSISNPELMIMHLANQTKSIKIGSGGIMPLHYSSFKLAETLKTLETCHPNRVS
SIGLGN SLGTVKVSNALRSLHKAHDYEEVLEELKSWLIDESSSKEPLVQPTLSSFP
DLYV LGSGQKSAYLAAKLGFTFGVFPMDKDPLTEAK

Sequence description:

A] Length: 588 bp - 196 aa (partial)

B] This gene sequence was not identified using the LEEP system. It was identified downstream of the ID-93 gene which was identified by LEEP, during cloning and sequence analysis of the full-length ID-93 gene sequence. No obvious signal peptide, but Shine Dalgarno sequence upstream of the ATG start codon.

FIG. 1_{CONT'D}

99 / 110

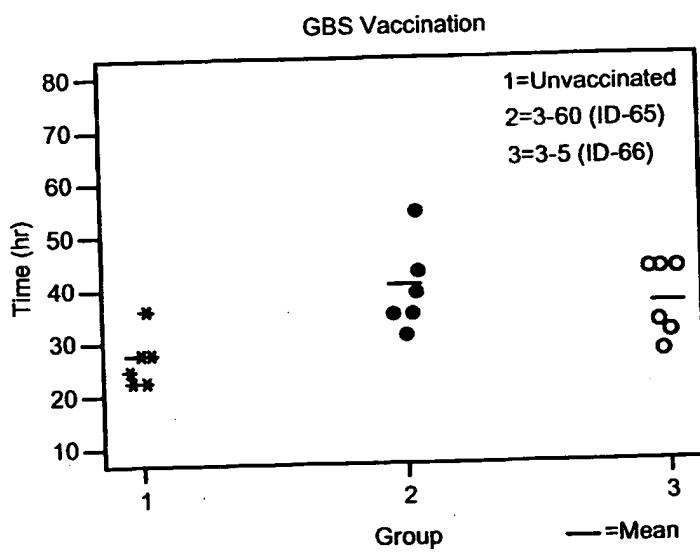


FIG. 2

100 / 110

nucS1

Bgl II Eco RV
5'-cgagattqatatctcacaaacagataacggcgtaatag -3'

nucS2

Bgl II Sma I
5'-gaagatcttcccggatcacaaacagataacggcgtaatag -3'

nucS3

Bgl II Eco RV
5'-cgagattqataccatcacaaacagataacggcgtaatag -3'

nucR

Bam HI
5'-cgggatccttatggacctgaatcagcgttgc -3'

NucSeq

5'-ggatgctttgttcagggtgtatc -3'

pTREP_F
5'-catgatatacggtacctcaagctcatatcattgtccggcaatggtgtggctttttgttttagcggataa
caatttcacac -3'

pTREP_R
5'-gcggatccccccgggttaattaatgtttaaacactagtcgaagatctcgcaatttcctgtgtgaaatt
gttatccgcta -3'

pUC_F

5'-cccaagggtttcccagtcacgac -3'

V_R

5'-tcaggggggcggagcctatg -3'

V₁

5'-tcgtatgtgtgaaatttg -3'

V₂

5'-tccggctcgatgtgtgaaattg -3'

FIG. 3

101 / 110

pTREP-Nuc vectors allow cloning of genomic DNA into each frame with respect to the nuclease gene

(i)

pTREP1-nuc1 (EcoRV) AAGTATCAGATCT--GATATC--TCACAAACAGATAACGGCGTAAAT Frame=+1
 ::::::::::::::: ↑ :::::::::::::::::::::
 ::::::::::::::: :::::::::::::::::::::
 pTREP1-nuc2 (Sma 1) AAGTATCAGATCTCCCCGGGA-TCACAAACAGATAACGGCGTAAAT Frame=+2
 ::::::::::::::: ↑ :::::::::::::::::::::
 ::::::::::::::: :::::::::::::::::::::
 pTREP1-nuc3 (EcoRV) AAGTATCAGATCT--GATATCCCATCACAAACAGATAACGGCGTAAAT Frame=+3
 ::::::::::::::: ↑ :::::::::::::::::::::
 ::::::::::::::: :::::::::::::::::::::
 TCACAAACAGATAACGGCGTAAAT

Nuclease Gene

Cloning site is indicated by an arrow

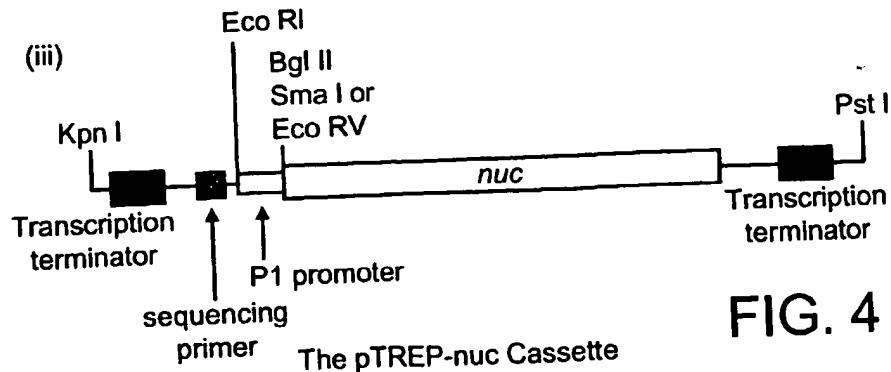
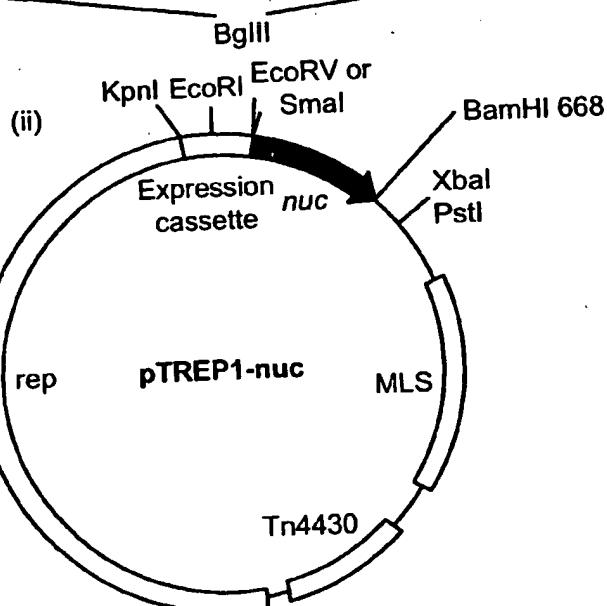
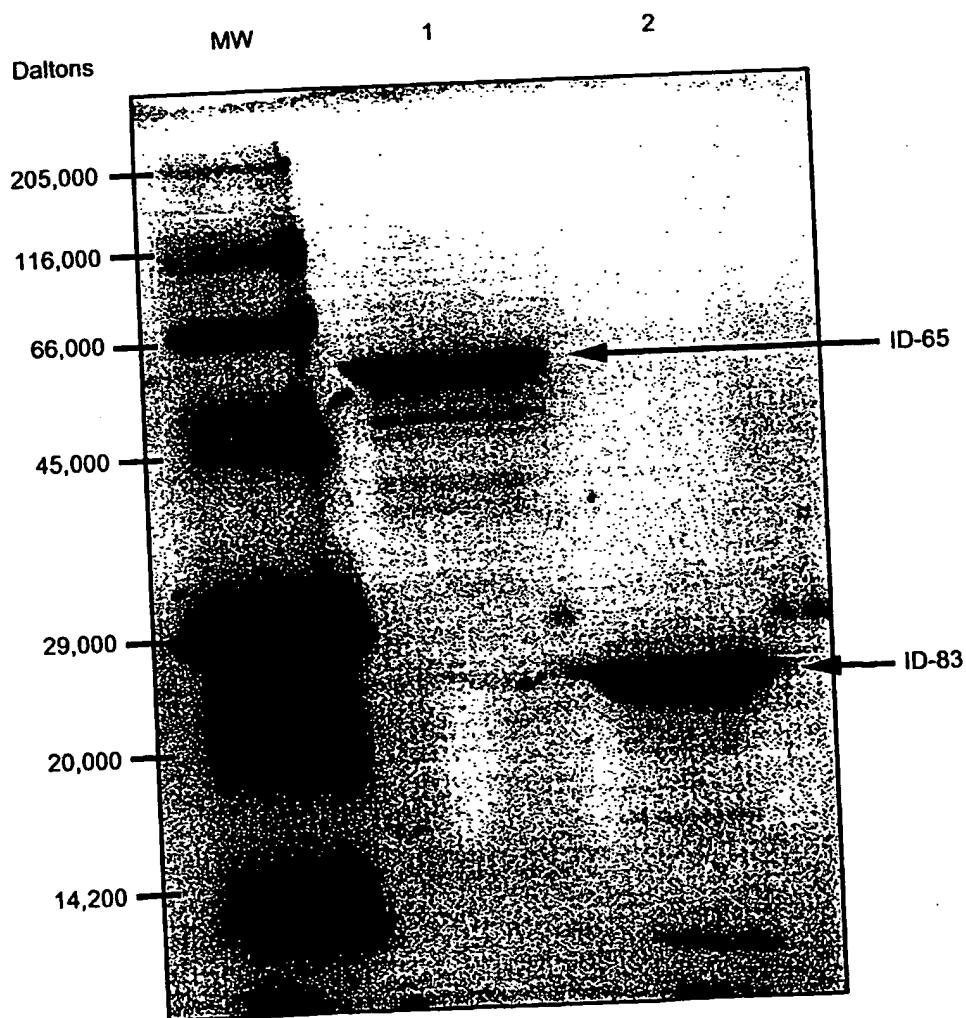


FIG. 4

102 / 110

FIG. 5

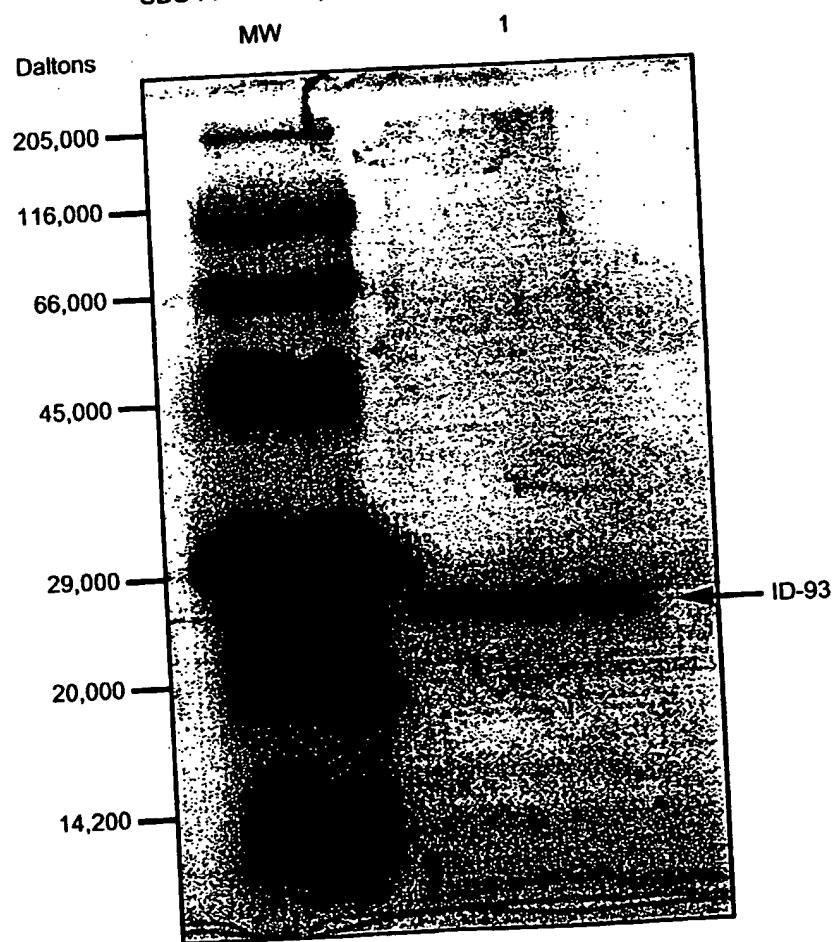
SDS-PAGE analysis of the purified ID-65 and ID-83 protein antigens



103 / 110

FIG. 6

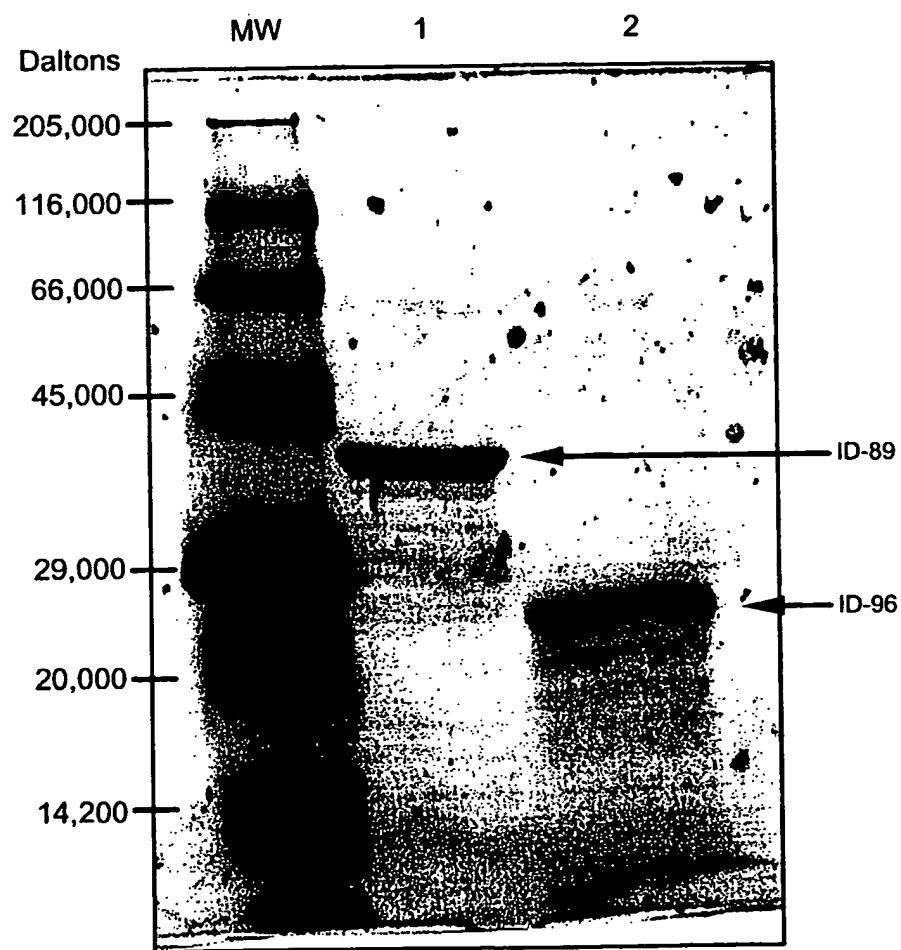
SDS-PAGE analysis of the purified ID-93 antigen



104 / 110

FIG. 7

SDS-PAGE analysis of the purified ID-89 and ID-96 protein antigens

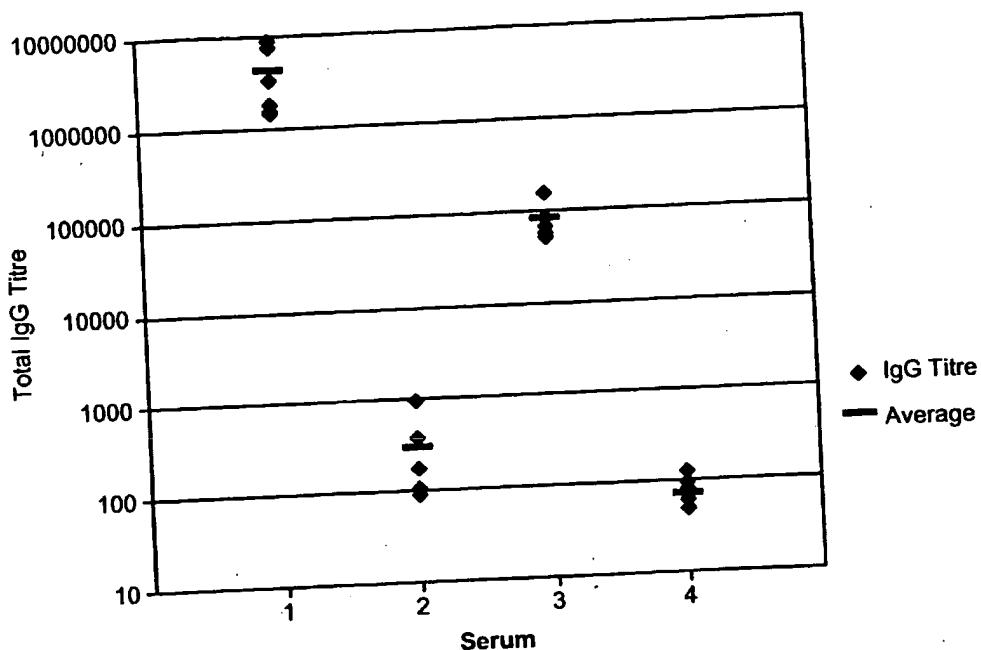


105 / 110

FIG. 8

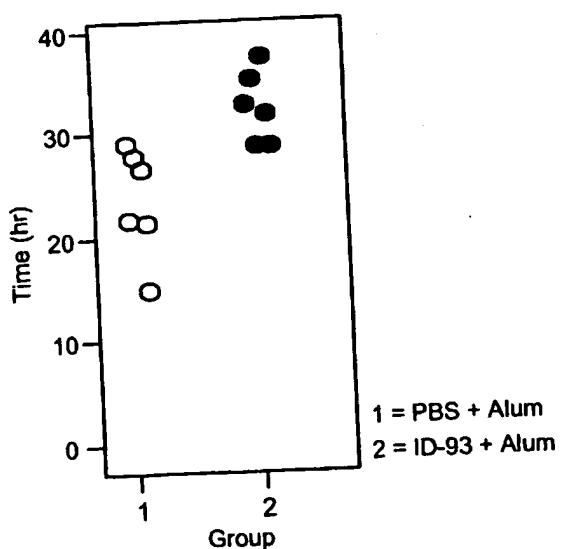
IgG Titres against the ID-65 and ID-83 proteins

ID-65 and ID-83 Vaccinations - IgG Titres

**FIG. 9**

Survival data

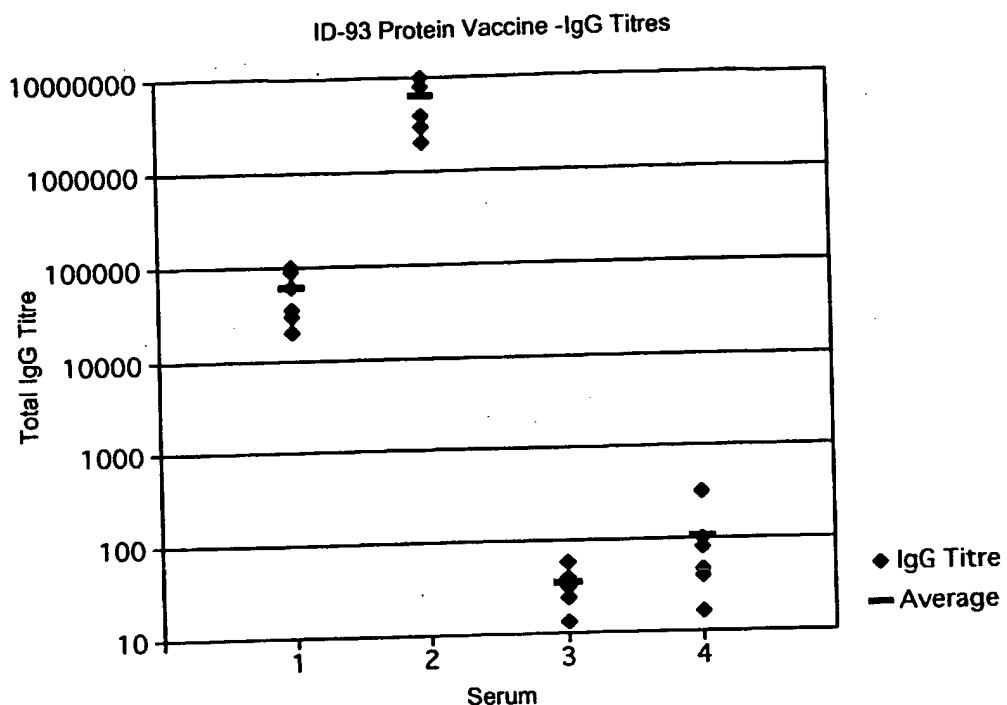
ID-93 Vaccination- GBS Challenge and Survival



106 / 110

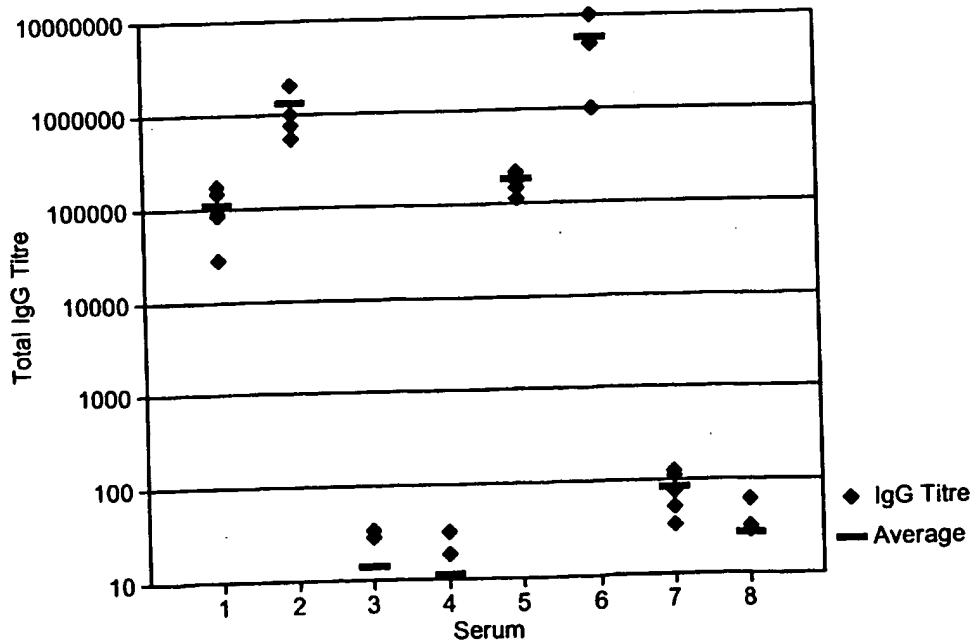
FIG. 10

IgG Titres against the ID-93 protein

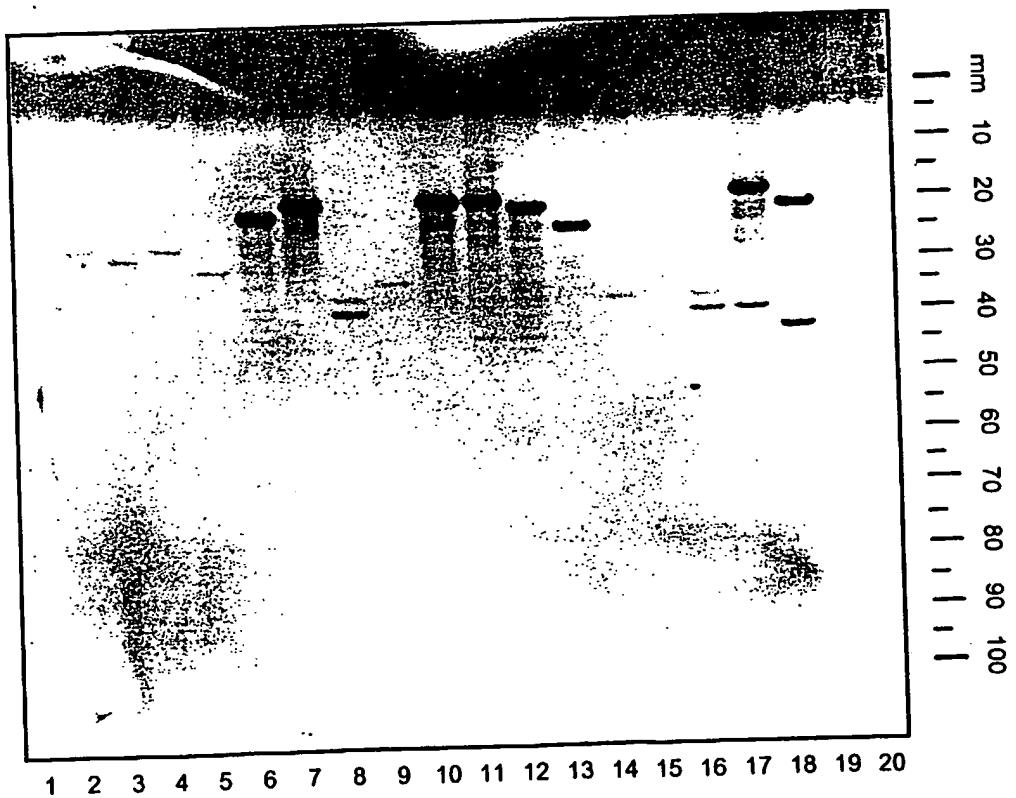
**FIG. 11**

IgG Titres against the ID-89 and ID-96 proteins

ID-89 and ID-96 Protein Vaccines - IgG Titres



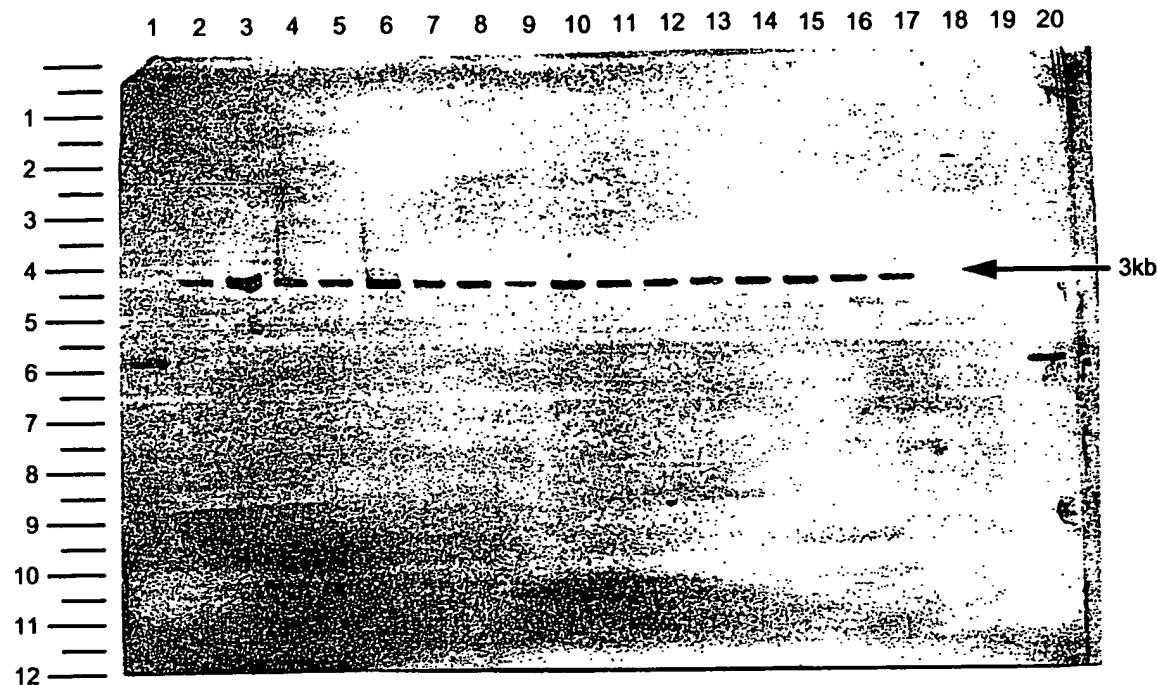
107 / 110

FIG. 12Southern blot analysis - *rib*

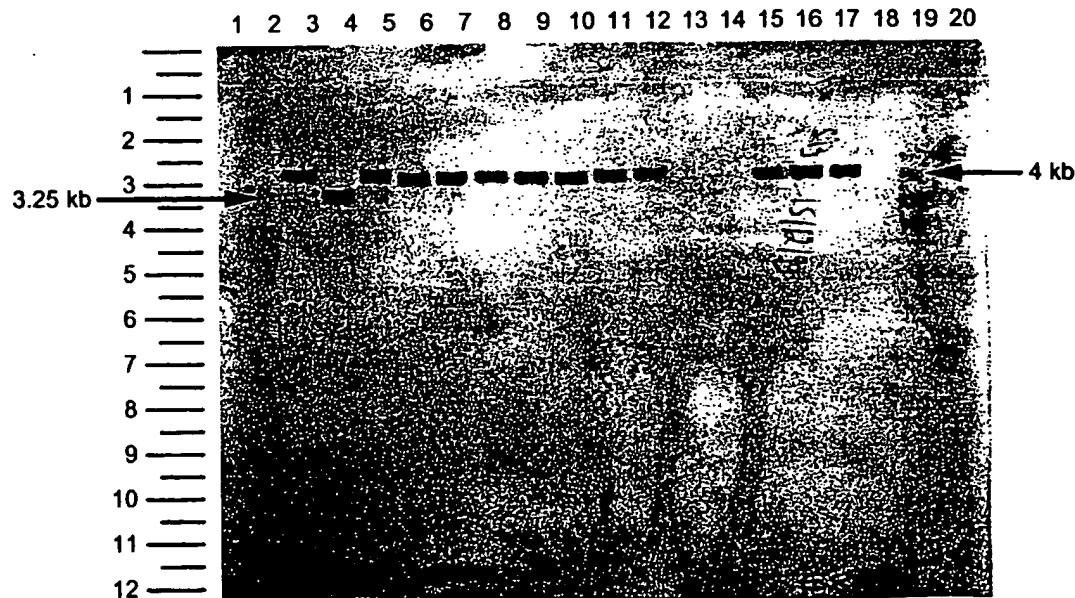
108 / 110

FIG. 13

Southern blot analysis - ID-65

**FIG. 14**

Southern blot analysis - ID-89

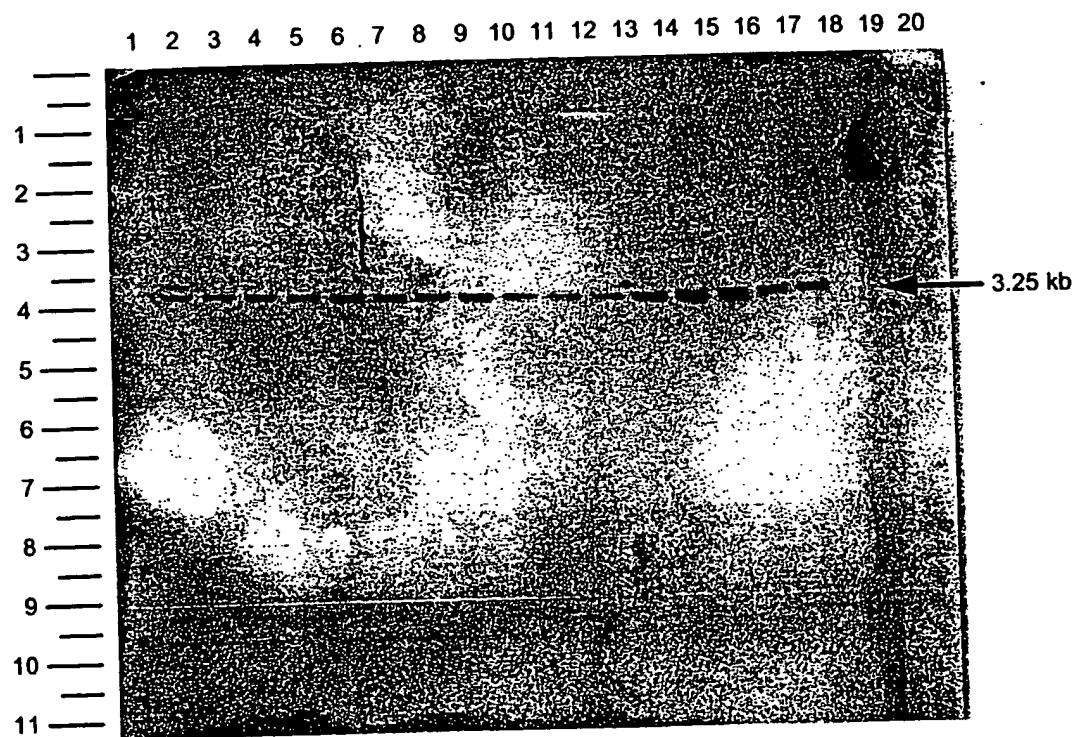


SUBSTITUTE SHEET (RULE 26)

109 / 110

FIG. 15

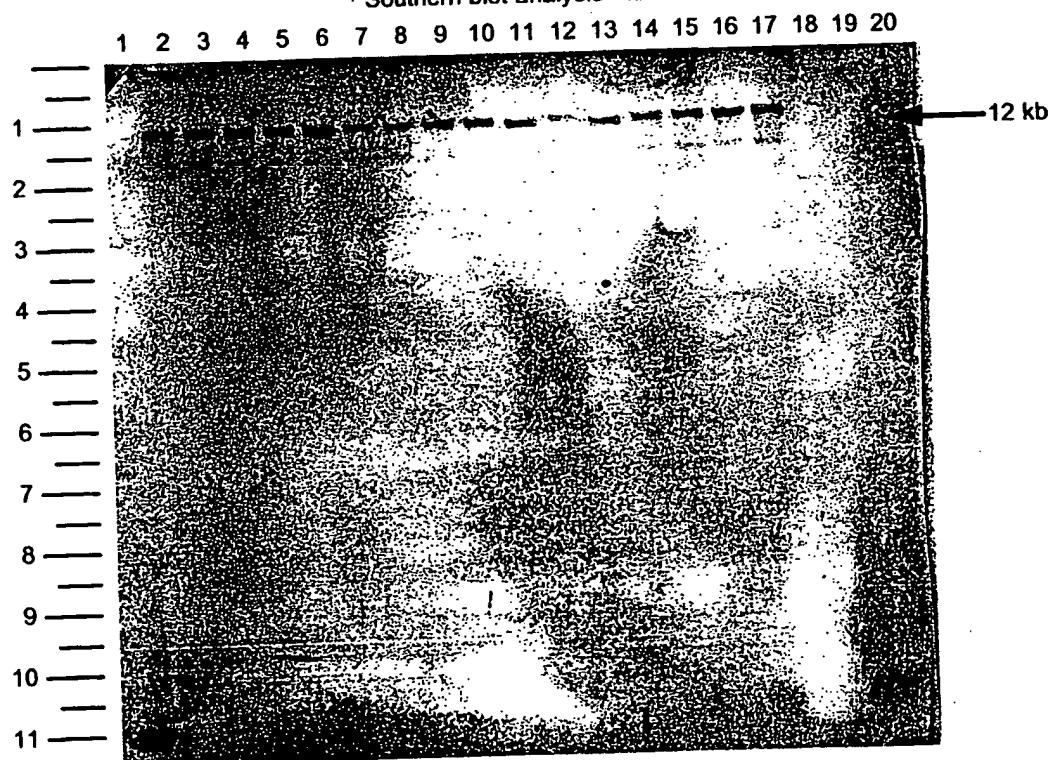
Southern blot analysis - ID-93



110 / 110

FIG. 16

Southern blot analysis - ID-96



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.